

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 05:57:59 ; Search time 2308.95 Seconds  
(without alignments)  
4322.143 Million cell updates/sec

Title: US-09-653-961-1

Perfect score: 1950  
Sequence: 1 atggggctccacagctgtg.....tgaggcattagcccggaat 1950

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba1:  
2: gb\_ba2:  
3: gb\_cm:  
4: gb\_ov:  
5: gb\_ph:  
6: gb\_pl1:  
7: gb\_pl2:  
8: gb\_pl2:  
9: gb\_pr2:  
10: gb\_pr3:  
11: gb\_ro:  
12: gb\_sy:  
13: gb\_sy:  
14: em\_fun:  
15: em\_hum1:  
16: em\_hum2:  
17: em\_in:  
18: em\_cm:  
19: em\_or:  
20: em\_ov:  
21: em\_pat:  
22: em\_ph:  
23: em\_pl:  
24: em\_ro:  
25: em\_sts:  
26: em\_sy:  
27: em\_un:  
28: em\_v1:  
29: gb\_da3:  
30: gb\_in1:  
31: gb\_in2:  
32: gb\_in3:  
33: gb\_pl3:  
34: gb\_pr4:  
35: em\_ba1:  
36: em\_ba2:  
37: em\_hc1:  
38: em\_hc2:  
39: em\_hc3:  
40: em\_hc4:  
41: em\_hc5:  
42: em\_hc6:  
43: em\_hc7:

44: em\_hc8:  
45: em\_hc9:  
46: em\_hc10:  
47: em\_hum3:  
48: em\_hum4:  
49: em\_hum5:  
50: em\_hum6:  
51: gb\_pr5:  
52: gb\_pr6:  
53: gb\_pr7:  
54: gb\_hc1:  
55: gb\_hc2:  
56: gb\_hc3:  
57: gb\_hc4:  
58: gb\_hc5:  
59: gb\_hc6:  
60: gb\_hc7:  
61: gb\_hc8:  
62: gb\_hc9:  
63: gb\_hc10:  
64: gb\_hc11:  
65: gb\_hc12:  
66: gb\_hc13:  
67: gb\_hc14:  
68: gb\_hc15:  
69: gb\_hc16:  
70: gb\_hc17:  
71: gb\_hc18:  
72: gb\_hc19:  
73: gb\_hc20:  
74: gb\_hc21:  
75: gb\_hc22:  
76: gb\_hc23:  
77: gb\_hc24:  
78: gb\_hc25:  
79: gb\_hc26:  
80: gb\_hc27:  
81: gb\_hc28:  
82: gb\_hc29:  
83: em\_hc30:  
84: gb\_hc31:  
85: gb\_hc32:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944.8	99.7	3335	10	AF089868 Homo sapi
2	1933.6	99.2	2943	85	HUMMUC18A
3	1933.6	99.2	3306	85	HUMMUC18B
4	1341.4	68.8	2106	11	AB035508
5	1333	68.4	2890	11	AB035508
6	1204.2	61.8	1988	11	AB035507
7	1197.8	61.4	2772	11	AB035509
8	275.4	14.1	2106	4	GHHEMCAMC
9	271.8	13.9	2584	4	CHK1GP
10	259.2	13.3	2015	53	HSWUC184
11	243.2	12.5	173038	76	HSWUC184
12	234	12.0	1998	4	GHHEMCAMC
13	232	11.9	2075	4	GHHEMCAMC
14	232	11.9	2365	4	CHKSGP
15	227.8	11.7	177595	76	AP001557
16	227.6	11.7	173038	76	AP001557
17	226	11.6	198902	76	AP001156
18	222.6	11.4	317	3	BTMCA1
19	222.2	11.4	267	11	MM0297451
20	213.8	11.0	270	53	HSWUC182
21	180	9.2	198902	76	AP001156

BEST AVAILABLE COPY

22	151.4	7.8	379	53	HSNKC185
23	151.4	7.8	177595	76	AP001557
24	136.8	7.0	6102	11	MMWGC18P
25	128.6	6.6	2225	11	MMWGC18G1
26	126.8	6.5	2882	53	HSNKC181
27	118.8	6.1	330	53	HSNKC187
28	117.4	6.0	141	52	HSJA297452
29	91.4	4.7	653	53	HSNKC183
30	87	4.5	141	11	MMU297450
31	55	2.8	2036	4	CHKASP
32	55	2.8	3361	4	GGBN
33	55	2.8	3381	4	S63276
34	52.4	2.7	1969	11	AFI09160
35	52.4	2.7	2402	11	AB035511
36	50.8	2.6	2356	11	AF221507
37	50.8	2.6	7218	81	T66494
38	48.6	2.5	2402	53	HSJLDBGRP
39	46.6	2.4	155074	8	AC003693
40	46.6	2.4	160847	65	AC026645
41	46	2.4	30041	29	SCJ9A
42	45.6	2.3	2059	11	AB035510
43	45.4	2.3	2396	53	HSBCK
44	45.4	2.3	2699	1	AF206633
45	45.4	2.3	37231	32	LMF1A803
					X68266 H.sapiens M
					AP001557 Homo sapi
					X74628 M.musculus
					X74627 M.musculus
					X68264 H.sapiens M
					X68270 H.sapiens M
					AJ297442 Homo sapi
					X68266 H.sapiens M
					AJ297450 Mus muscu
					M76678 Chicken Ig
					X64301 G.gallus mR
					S63276 SCI-100 kda
					AFI09160 Mus muscu
					AB035511 Mus muscu
					AF221507 Mus muscu
					166494 Sequence 14
					X83425 H.sapiens L
					AC003693 Human Chr
					AC026645 Homo sapi
					AL109972 Streptomy
					AB035510 Rattus no
					X80026 H.sapiens B
					AF206633 Streptomy
					AL161416 Teishman1

## ALIGNMENTS

RESULT	1
LOCUS	AF089868
DEFINITION	AF089868 3335 bp mRNA
VERSION	AF089868
KEYWORDS	complete cds.
SOURCE	AF089868.1 GI:4336423
ORGANISM	human.
REFERENCE	1. Homo sapiens
AUTHORS	Eukaryotic: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL	2. (bases 1 to 3335)
REFERENCE	Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
AUTHORS	PIH12 from human umbilical vein endothelial cells
TITLE	Unpublished
JOURNAL	2. (bases 1 to 3335)
REFERENCE	Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
AUTHORS	Direct Submission
TITLE	UMC14 (02-SEP-1998) Medicine, University of Minnesota, Box 480
JOURNAL	Sumitted (02-SEP-1998) Medicine, University of Minnesota, Box 480
FEATURES	Location/Qualifiers
source	1..3335
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_type="umbilical vein endothelial"
	/note="Primary culture"
	24..1964
	/note="PIH12 antigen"
	/codon_start=1
	/product="cell surface glycoprotein PIH12 precursor"
	/protein_id="AADI17799.1"
	/db_xref="GI:4336424"
	/translation="MGDLRYCAFLAACCCPRVAGVPGAEADQPAPELVEVSGSTA
	LKGLCSOSGNSLHWDFESVHKKEKRLTIFVRGOGSGEGEYERLSTLDKQATLTA
	LTQVYQDERFLFCGKKRPSQEQEIRLTIRYKAPFEPNIGVPLGIPVNSKEEVAT
	CGRGNGYPIPOVIWIKNGRPLKEEKNRVHDSQTSVSSGLITQSLIKQVLEKDD
	AGPELRYRLPSPGNHMKSESEVYTPVPEYKELVEYEPGMKLKEDRYEIRLADG
	NPPPFISKONPSTREAEEETITDNGVLTPARKHSGRYECQGLDPTMISLSE
	PEILLVNVSDVRSVPAPEREQESLTITCEASDLEFQMLREETDQVLEPVL
	QHLHDKREAGGGYRCVAVSPISGLRTQLVNVAFIIGPNNAFKERVAKENNVNL
	SEASGSHPRPTSMKNVNGTSAPDQDDPQSVLTSLVLTPELTGVECTASDLGNK
	TSITLAEIVNLTTLTPDSTSTTGSLSTSTASVTRANSSTERKLPEESRGCVIAVAT
	VTILVAVAGAVLYFLYKNGKRLPCRSRSGKQETILPEPSKSLVYVSDKLPPEEMGIL

[illegible]

LOCUS	DEFINITION	HUMANUC18A	2943 bp	mrna	PRI	16-AUG-1994
QY	961	caggccctcgagacttgagacacacatgatactgctcgtctagtgtaaccacaggaactctacggtg				1020
Db	984	CAGGCGCTCGACTTGGACACCACTGATATTCCTGCTGATGTAACCAACAGAACTACTGtg				1043
QY	1021	aactcgtgctcagcgtccgagtgagtcgcccgagccctctagagacagagaagtcagcagc				1080
Db	1044	AACATATGTGTGACGTCGCGAGTgATGTCGCGAGCCCTCTAGAGACAGAAAGGCACTACG				1103
QY	1081	ctcaacctgacctgtgtgagcaagagatagccaagacctctgagttccaagtgctgtagagaa				1140
Db	1104	CTCACCTCGACCTGTGAGGCGAGAGTAGAGACAGACTCGAGTTCCAGTGGCTGAGAGAA				1163
QY	1141	gagacagcgccaggtgctcgtggaagaaggggcccgtgctctaagttcatgatgacctgaaacggag				1200
Db	1164	GAGACACACAGGtGTGGAAAGGGGGGCTGTGCTTCAGTTGCATGACCTGAAACGGGAG				1223
QY	1201	gcaagagcgagctctcgtcgtgagcgctctgtgtcccaagcatacccgagctcaacgcacaa				1260
Db	1224	GCAGGAGGCGGCTATCCCTGCGTgGCTCTGTGCCCCAGCATACCCGGGCTGAAACGGCACTA				1283
QY	1261	caagctgtgtcaacgctgtgcccatttttggcccccttgatgtgcatctcaagagagaagagtg				1320
Db	1284	CAGCTGTGCAACGtGGCCATTTTGGCCCCCTTGATGTGCATTCAGAGAGAGAGAGTg				1343
QY	1321	tggtgtgaaagagataatgtgtgttgatctgtctgtgaaagctcagggcaacccccggccc				1380
Db	1344	TGGTGTGAAGAGAAATATGTGGTTGAATCTTCTTGTGAACGTCGAGGGCACCCCGGCC				1403
QY	1381	acctctcccttgaaagctcaagcgcaagcggaagtgaacaagaaccaagatccacaagcgagtc				1440
Db	1404	ACCATCTCTGTGAACGTCMAAGGCGACGGCAAGTGAACAAACCAAGATCCACAGCCAGTTC				1463
QY	1441	ctgagcaacctgaaatgtcctcgttgaccggcgagcgtgtgtgagacaggtgtgtgaatgacg				1500
Db	1464	CTGAGCAACCTGAAATGTCTGTCGTGACCCCCGAGGCTGTTGAGAGACAGGTGTTGAATCAAG				1523
QY	1501	gctctcaagagcctcgtgggcaaaaacaacagcatcctcttccctgtagctgtgtcaatlaacc				1560
Db	1524	GCCTCCAAACGACCTGTGGGCAAAAACACAGCATCCTCTTCTGTGAAGCTGTGCAATTTAAAC				1583
QY	1561	acctctcaacagagctccaacaacaacacgctcgtctgaagccttcacatgcctgcaatctccat				1620
Db	1584	ACCCTCACACCACTCCACACACACACTGgCCTTCAGCACTTCCACTGCAAGTCTCTCAT				1643
QY	1621	accagagaccacaacagcacctccacaagagagaagaagctgtccgagacccgagacccggcgctg				1680
Db	1644	ACCAAGAGCCAAACAGCACACTCCACAGAGAGAAAGTGGCGGAGCGCGAGACCGGGGGGTG				1703
QY	1681	gtcaatcgtgtgctgtgtatgtgtgtcatccctgtccctgtcctgtgagtgctgtgagcgtctccat				1740
Db	1704	GTCATCGTGGCTGTGATTTGTGTGTCATCTGTGCTCTGCGGGTGGCGCTGTCTCTCAT				1763
QY	1741	ttctctcataaagaggaagctctccgtctgaagcgctccaaggaagcagagatccaagctg				1800
Db	1764	TTCTCTTATTAAGAAAGGCAAGCTGCCGTGAGGGGCTCAAGGAAAGCAGGAGATCAGCTG				1823
QY	1801	cccccgctctcgtlaagagcggaactgtgtgtgaagttlaagtcagaataagctccccaagaag				1860
Db	1824	CCCCGCTGCTGCTAAGAGCGAACTGTAGTTGAAGTTAAGTCACATTAAGCTCCACGAAGAG				1883
QY	1861	atggagcctcctcgtcagggcaagcagcggtgacaaagaagggctcccggaagacaaggaagaaa				1920
Db	1884	ATGGGCGCTCTCGTCAGGGGACGACGCGGTGACAAAGAGGGCTCCGGGAGACCCAGGAGAGAAA				1943
QY	1921	tacatgatctgagacattagccccgaa				1948
Db	1944	TTCATTCGATCTGAGGCAATTAGCCCCGAA				1971

ACCESSION	CDS.
VERSION	M29277
KEYWORDS	M29277.1 GI:530047
SOURCE	MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like sequence; integral membrane glycoprotein.
ORGANISM	Human cDNA to mRNA.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 2943)
JOURNAL	Lehmann,J.M., Riettmuller,G. and Johnson,J.P.
MEDLINE	MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)
TITLE	2 (bases 1 to 2943)
JOURNAL	Sers,C., Kirsch,K., Rothbacher,U., Riettmuller,G. and Johnson,J.P.
MEDLINE	Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains
AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
REFERENCE	9339184
AUTHORS	3 (bases 1 to 2943)
TITLE	Johnson,J.P.
JOURNAL	Direct Submission
FEATURES	Submitted (06-Oct-1989) Judith P. Johnson, Institute for Immunology, University of Munich, Goethestrasse 31, Munich, Germany 80336
COMMENT	On Aug 16, 1994 this sequence version replaced gi:188859.
FEATURES	Location/Qualifiers
SOURCE	1..2943
	/organism="Homo sapiens"
	/isolate="Juso"
	/db_xref="taxon:9606"
	/tissue_type="melanoma"
	/cell_line="Mel Juso"
	/clone="drop4.7"
CDS	8..1948
	/codon_start=1
	/product="MUC18 glycoprotein"
	/protein_id="AA20824.1"
	/db_xref="GI:530048"
	/translation="WGLPRLVCAFLAACCCCPVAGVPCGABAPAEIYVEYGSTRA LKQGLSQSGNLSHVDMFVSFKRKRLIFRRVGQSGSEGEYQRSLSDRCATATA LTQVTPDERFLFCQGRPRSOEYRIOLRYKAPEENINIVNPGLGVNSKEPEVAT CVGNNGYPIDOVIVYKNRPLKEKNKHIOSSOTVSSGYLTSSLKOLVEDKD AOFCEINLRPLPSGNHKMKESRETVVPVFYEPEKMLFEVPGMKLEDREIRCLAG NPPHEFSIKONPSTRERAEETITDNKVILLEPARKKSHGLEQOANNIDTMSLISE PQELIVSVSDVRVSPARPERQEGSSLTFLTEAESSODLEROWMERBTDVLIEGPVL QHLDRKEAGGVCACVASVPSIPGLNTOLVKLAIFCPPMMAFERKRVAKENVAVNL SCENSGRPRTISWNVNGTASEODOPORVYSTLVNVLPFELLETTGECASNDLGNC TSIILEFVNATLTLPDSENTTGTSASTPHTANSTERRKLDPESGGVIVAVIT VCTIIIVLAGVILVELTKKGKLPKRSGKEITLPPSRKTEIIVEVKSDDLPEMGILL OGSSDKRARREDQSEKYTLDRH"
	8..91
sig_peptide	/product="MUC18 glycoprotein signal peptide"
mat_peptide	92..1945
	/product="MUC18 glycoprotein"
misc_feature	122..373
	/note="immunoglobulin-like (V set) domain I"
misc_feature	467..694
	/note="immunoglobulin-like (V set) domain II"
misc_feature	795..988
	/note="immunoglobulin-like (C2 set) domain"
misc_feature	1076..1249
	/note="immunoglobulin-like (C2 set) domain"
misc_feature	1337..1525
	/note="immunoglobulin-like (C2 set) domain"
BASE COUNT	668 a 863 c 858 g 554 t
ORIGIN	
Query Match	99.2%; Score 1933.6; DB 85; Length 2943;





Db	1028	AACTATGTCTGACGCTCCGAGTGAATCCCGACGCCCTTGAGAGACAGCAAGCAGCAGC	1087
Qy	1081	ccacccctgaactctgtagagcagagtagccagagccctgagcttcacgtggtctgagagaa	1140
Db	1088	CTACCCCTACCTCTGTAGGACAGAGAGTACCCAGACCTTCAGATTCCAGTGGCTGAGAGAA	1147
Qy	1141	gaagacgagcaggtctctgtagaaggggacctgctcagctcagctcagctcagcagggag	1200
Db	1148	GAGACAGACAGAGTGTCTGGAAGGGGGCTGTGCTTCAGTTCAGTACAGCTGAAACGGGAG	1207
Qy	1201	gcagagagcggtactatcgtctgctgctgctctgctgctgctgctgctgctgctgctgctg	1260
Db	1208	GCAGGAGGGGGGCTATCGCTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1267
Qy	1261	cagctgtagaagtgagcatttttggcccccttgagtagcattcaaggaaggaaggtg	1320
Db	1268	CAGCTGTGAAGCTGGCCATTTTGGCCCCCTTGATGATGCAATCAAGAGAGAGAGGTG	1327
Qy	1321	tggtgtagaagagatactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1380
Db	1328	TGGGTAAAGAGAAATATGCTGTTGAAATCTGCTTGTGAAGCCTCAGGCGACCCCGGCC	1387
Qy	1381	accatctctggaacgtcaacgagcagtgaaatgaacaaagaatccacagcagatc	1440
Db	1388	ACCATCTCTGGAAGCTCAACGCGCAGCAAGTGAACCAAGACCAAGATCCACAGCGAGTC	1447
Qy	1441	ctgagacacctgaatgctctctgtagaccccgagctgctgtagaagcaggtgtgaatgacag	1500
Db	1448	CTGAGACCCCTGAATGCTCTCTGAGACCCCGAGCTGTGTGAACAGGTGTTAAATGACAG	1507
Qy	1501	gctccaaagacccctgtagaagaaac	1560
Db	1508	GCTTCAAGAGAGCTGGGCAAAAAACACACACATCTCTCTGAGAGCTGCTCAATTTAAAC	1567
Qy	1561	accctcaacacagac	1620
Db	1568	ACCCTCAACACAGACTTCCAAAC	1627
Qy	1621	accagagcacaagcagcctccacagagagaagacctgctgagagccgagagccgagagcgctg	1680
Db	1628	ACCAGAGCCACAGCACCTCCACAGAGAGAAACCTGCCGAGCCGAGAGCGGAGCGGCGTG	1687
Qy	1681	gtcatctgtagctgtagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct	1740
Db	1688	GTCATCTGTGGCTGTGATTTGTGTCATCTGTGCTGCGGTGGCTGCGCTGCTGCTGCTGCTG	1747
Qy	1741	ctccctcataaagaagcagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1800
Db	1748	TTCCTCTATAGAAAGGCAAGCTGCGCTGCAAGCGCTCAGGGAAGCAGATACACGCTG	1807
Qy	1801	ccccgctctcgtgaagcgaactgtgtagctgtagctgtagctgtagctgtagctgtagctgtag	1860
Db	1808	CCCCGCTCTGTAAGACCGAATCTTGTAGTTGAATTAAGTCAAGTATAGCTCCACAGAGAG	1867
Qy	1861	atggagctctctgtagcagcagcagctgtagcagcagcagcagcagcagcagcagcagcagcag	1920
Db	1868	ATGGGCTCTCTGCAAGGAGCAGCAGCGGTGACAAAGAGGGCTCCGGAGACACAGGAGAGAA	1927
Qy	1921	tacatcgatctgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	1948
Db	1928	TACATCGATCTGAGGAGATTAGCCCGAA	1955

RESULT 4  
AB035506 2106 bp mRNA ROD 03-OCT-2000  
DEFINITION Rattus norvegicus mRNA for 1-glicerol/MUC18, complete cds.  
ACCESSION AB035506  
VERSION AB035506.1 GI:10566948  
KEYWORDS 1-glicerol/MUC18.  
SOURCE Rattus norvegicus (strain: Sprague-Dawley) male CDNA to mRNA.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	1 (bases 1 to 2106) Taira,E., Okumura,S. and Miki,N.	
JOURNAL	Functional analysis of rat glicerol/MUC18	
TITLE	Published Only in Database (2000) In press	
REFERENCE	2 (bases 1 to 2106) Taira,E., Okumura,S. and Miki,N.	
AUTHORS	Direct Submission	
JOURNAL	Submitted (03-DEC-1999) to the DDBJ/EMBL/Genbank databases. Eiichi Taira, Osaka University Medical School, Department of Pharmacology, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp, Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)	
FEATURES	Location/Qualifiers	
source	1..2106 /organism="Rattus norvegicus" /strain="Sprague-Dawley" /db_xref="taxon:10116" /sex="male" 17..1963 /gene="1-glicerol/MUC18" 17..1963 /gene="1-glicerol/MUC18" /product="1-glicerol/MUC18" /protein_id="BAB16048.1" /db_xref="GI:10566949" /translation="MGILPVLVCFLEAACCCSANTGVGEKQPTPTPDVPEVNGN TALLKCPAPHSNFSQVWELFLHKEKQIDIFVHOGKQSEGEYEHRLSLHPEAT LALSOVPHDRLKSKOPRPODHYVOLQVYKAEPTIQANVGIHVDIOELKEV ATCGVNGRYPIPOVIMYKNRPLQEEENRPILOSQTVSSGYTLTKSVLSARVYKED KDOFCEVSEYRPPSGNRKSEKVTVPVLYPAEKVVEVPEGLKLESDHYKICILT DGNPOPHFTINKNPSEMEKEESTDENGILSLPEAKHHSVGYOOSLDLEFTWLS SDPELLVNVSVQVDPTRAPRYQEDDSLITCKASNDLREMTIRDKTGGLLKGP ILQINNVKREAGRYLCAVSPVPGALNRRVSVGIFSPMAKERKVMQENML NLCEASGHQPTISWNIKGSATENMPDQVTVSTLVNLTPELLETGAECTANSLG SYTTVILKLVLTTLTHDSQYTGSLTPVSPHSRANSTETKLPQDSGQVYIVA VICTVLAVLAVGATLYFYFKGKLPGRSGKQKITLPTPRKSEFVEVKSDBLPBEA LLGNSNDRKAPDDGKXYIDLRH"	
BASE COUNT	536 a	573 c 587 g 410 t
ORIGIN		
Query Match	68.8%;	Score 1341.4; DB 11; Length 2106;
Best Local Similarity	80.9%;	Pred. No. 9.8e-282;
Matches 1577; Conservative	0;	Mismatches 366; Indels 6; Gaps 1;
Qy	1 atggagctccagcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	60
Db	17 ATGGGCTGCCCCAGGCTGTGCTGCGCTTATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	76
Qy	61 gtcgaggttgcccgagagcctgagagc-----cctgagcctgagcctgctgtagagtg	114
Db	77 GCCACGGGTGGCCAGGAGAGAAAGCAAGCCACACCCAGCCGAGCCGGTGAAGTA	136
Qy	115 gaagtggcagcagccctctgtagctgagcctctccagctcccaaggaacacactcagc	174
Db	137 GAAGTGGCACAAGCGCCCTCTGAGAGTGCGCCCGACATCCCTCAGGAACCTTACG	196
Qy	175 catgtgactgtttctctgccaagaagagacgagcgtcatctctcgtgtgagccag	234
Db	197 CAAGTGAATGATGTTTGTATTCACAGAGCAGCAGATATCCCATTTTCCGTTGACACAG	256
Qy	235 ggcagagcagagcagcagcctgtagagctgtagagcagcagcctccaggaagagag	294
Db	257 GGCANAGGCCAGATGACCTTGGGATATGAAACCGCTCAGCTTCCACGGCCCAAGGA	316
Qy	295 gctactctgcccctgactcaagtaacccccaagacagcagcactctctgtgtgccaagc	354
Db	317 GCTACTCTGGGCTGAGTCAGAGTCACCTCCCATATGACCGGATGTTCTGTGTAAGAGC	376
Qy	355 aagcgccctggttccagagtagccagcagcagcagcagcagcagcagcagcagcagcagcag	414

```

Db 377 AAGCAGCCACGGCCCGCAGGATCATGTTCAGCTGCAAGTCTACAAGGCTCCAGAGAG 436
Qy 415 ccaaacatccagtgtaaacccccctgggcatccctgtgaagaatgaagagctgaagatc 474
Db 437 CCAACTATTCAGGACGAGTGGTGGGCAATCCACGTGAGATCAAGAGTTAAGAGGTT 496
Qy 475 gctacctgtgaaggaggaacggtgtaacccatctctcaatgcatctggtacaagaatgac 534
Db 497 GCTACTCTGTGGGAGGAGATGCTTACCCTTCTCAAGTATCTGTGTACAAAGATGGC 556
Qy 555 cggcctctgaaggaggaagaacccgggtccacatcagtcgltccacagctgtgagtcg 594
Db 557 CGGCCCTTGCAAGAGAGAGAGAACCGCGTTTCATATCCAGTCTGCACAGACCGTGGAGTCC 616
Qy 595 agtggttgttaacctctgacagatctctgaaggacacagctgtgttaagaagaagaat 654
Db 617 AGTGCTTGTACCTTGAAGAGTGTCTTAAGTGCACGGGTAGTCAAGGACACAAAGAT 676
Qy 655 gccagtttactgtgagctcaactaccgctgcagctgagtggaacacatagaagagtc 714
Db 677 GCCCAGTTTACTGTGAGTCAAGTCAAGCCGCCGCGCCAGTGGGAGACCGCTGAAGATCT 736
Qy 715 agggaagtcacccgtctcgtgttttctaccgacagaaagaatgtgtgctggaagtcgccc 774
Db 737 AAGGAGGTCACTGCTCCCTGTATTATCCCTGCAGAAAAAGTGTGGGTGAGGTAGAACCC 796
Qy 775 gtgggaagtcggaaggagggagccgctgtggaatcaagtggtgttgctgtgacaccc 834
Db 797 GTGGGAGTCTGAAGGAGGAGGACCATGTAAGATCAAGTGTCTCACTACCGCAACCT 856
Qy 835 ccaccacactcaagcatcagcaagacagacccacagacacaggaaggaagaagaagaca 894
Db 857 CAACCCCACTTACATATCAACAAGAAAGAACCCACAGCATGAGAGATGGAAGAAGAGAC 916
Qy 885 accaagcaacaggggtccctgtgtgtgagcctgcgcggaagaaacagtggtgcgtat 954
Db 917 ACCGATGAAAAAGGCTCTCTCTCTTGGAGCCGCCAGAAACACACAGTGGGCTCTAC 976
Qy 955 gaatcagggcctgtgacacacatgatcgtcgtgtgagtggaacacagaacta 1014
Db 977 CAATCTCAGAGCTTAGACTTGGAAACTACAGTTATGCTGTCAAGTACCCCTTGGAGCTG 1036
Qy 1015 ctgtgaaactatgtctgacgctccgagtgagtcgagcccgccctgagaagaagaagc 1074
Db 1037 CTGGTAACATATGTCTGTGATTTCAAGTGAATCAACCTGCCCTGTAATCCAGAGAT 1096
Qy 1075 agcagcctacccctgacctgtgagcagagtagcagagacctcgagttccagtgctg 1134
Db 1097 GATACCTTCACACTGACTGTAAAGCAGAGATACACAGGACCTTGAGTTGAGTGGCTG 1156
Qy 1135 agaagaagagagagcagagtggtggaaggggagcgtgtgctcaatgcatgactgaaa 1194
Db 1157 AGAGCAAGAGACGGCCAGCTTGTGGAAAAGGGCCCATCTTCACTTAACAACGTGAAA 1216
Qy 1195 cggagagcaggaagggagctatcgtcgtgctgtgtgtgccaagatacccgagctaac 1254
Db 1217 AGGGAAGAGAGGGGGAAGATATCTGTGTGATGTGTGCCATGTGCCAGGCTTCTGCTTGAAT 1276
Qy 1255 cgcacagcctgtgtacaagctttgtgcccccttgatgacatcaacagagagag 1314
Db 1277 CGCACACGCGGGTCTAGCTGGGCAATTTTGGTCCCGTGTGATGCGACGAAGAAGAGAG 1336
Qy 1315 aaggtgtgtggaagaataatgtgtgtgatactgtctgtggaagcgtcagggcacc 1374
Db 1337 AAGGTGTGGGCGCAAGAAATGCTGAATGTCTTGTGAAGCTTCAAGGCCATCCT 1396
Qy 1375 cggccaccatctctgtgaacgltcaacgcaagcaagtgaaacaagaatcacacag 1434
Db 1397 CAGCCACACATCTCTGTGAATATCAATGCTGTGGCACTGAATGGAACCTGTCTCTCAG 1456
Qy 1435 cgaagtcctgaagcactgaatgtctcgtgaccccgagcgtgtgtgagagagtgltgaa 1494

```

```

Db 1457 ACAGTAGTAGACACTTGAATGTCCTTGTGACCCAGAGCTTGTGGAGACCGGTGCAGAA 1516
Qy 1495 tgaaggtctccaagaagcctgtggcacaacaccacagcatctctctcgtgagctgtgcaat 1554
Db 1517 TGTACAGCTCCACATCCCTGTGGGCTTATACACAGCTCATCTATCTGTAAGCTGTCACT 1576
Qy 1555 ttaacacccctacacagactccaacacacacacacacacacacacacacacacacacac 1614
Db 1577 TTATACCATCTTACATCATATCTTACACCCACACACTGCGCTTGTGACACCCCAAGTCACT 1636
Qy 1615 cctcatlacagagcaacagcactccacacagagaagaactgtgcggaagccggaagccg 1674
Db 1637 CCTCACAGAGAGCCACAGACCACTCTCCACAGAGAAAACCTGCCACAGCAGAGAGCAAA 1696
Qy 1675 ggcgtgtcatcgtgtgctgtgattgtgtgcatcctgtgtcctgtgcgtgtgtgtgtc 1734
Db 1697 GGTGTGCTATGCTGTGCTGTGATGATATGATGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1756
Qy 1735 ctctatctctctataagaagagcaagctgtgcgtgtgagcgtctcaggaagcagagatc 1794
Db 1757 CTCTATTAATCTTACAGAGAGGCAATGCTCTGTGACGCTCAGAGAAACAGAGATC 1816
Qy 1795 acgtgtccccgtctcgtlaagaagcgaactgtgagtgaagttcaagtlcaagatccca 1854
Db 1817 ACGTGTCCCGCCAGCTGTGAGAGTGAATTTGATTGAAGTTAAGTACAGATTAAGCTCCA 1876
Qy 1855 gaagagatgtgctcctcgtgagggcagcagcgtgtgacaagaaggtctccggaagcagga 1914
Db 1877 GAAGAGATGCTCTCTCTTGAAGGACAGCAAGGTGACAAAGAGGCTCCAGAGACAGGGA 1936
Qy 1915 gagaatcatcgaatctgagcatgagc 1943
Db 1937 GAGAAATACATGATCTGAGGCAATTAAGC 1965

RESULT 5
AB035508
LOCUS AB035508 2890 bp mRNA ROD 03-OCT-2000
DEFINITION Mus musculus mRNA for 1-glicerol/MUC18, complete cds.
ACCESSION AB035508.1 GI:10566952
VERSION 1-glicerol/MUC18.
KEYWORDS Mus musculus (strain: BALB/c) male cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2890)
AUTHORS Taira, E., Okumura, S. and Miki, N.
TITLE mouse glicerol/MUC18
JOURNAL Published Only in Database (2000) In press
AUTHORS 2 (bases 1 to 2890)
Taira, E., Okumura, S. and Miki, N.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases. Eiichi Taira, Osaka University Medical School, Department of Pharmacology; 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp, tel: +81-6-6879-3521, fax: +81-6-6879-3521)
FEATURES
source
1..2890
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
20..1966
/gene="1-glicerol/MUC18"
20..1966
/gene="1-glicerol/MUC18"
20..1966
/codon_start=1
/product="1-glicerol/MUC18"
/protein_id="BAB16050.1"
/db_xref="GI:10566953"
/translation="MGIPKIVCFLFAACCCBRAGVPEEKOPVPTDVEAEVGS

```

TALLKCPBRASACNFSOYDWMFLIKKEQOLIFRPHQCKQAREPEYEHRLSLQDSVAT  
LALSHVTHPEDEMLCKSKREPILODHDYVLEQVFAPEEPTIOANVVSIVHRODLREV  
ATCVGRNGYPIPOVLWYKNSLPILOEENRHYVISOSSGLTTEKSYLSARLAKED  
KDAQFCELSYRLPSGHNHKESEKVEVYVPEPAKWVWEVAPVLLEQGHVITRCLT  
DNQNPQHTLKKKDPSTGEMEEESTDNGSLIPAPAKHNHSGVICOOSIDLEPTITLS  
SDPELLVNTYSYDVQVNPPTAPEVQDEBSLTLTCAESNODLEFMYLDRKQGLCKRP  
VLQLONNVRRRAGRGYRLCMASVPRVQGNRLQVLSVGLFGSMVILKRWKYNQDANV  
NLSCASGHPQPTISMWNVGSAETBMPNDPOVUWLVNVLPIPELLETGACDTSNSJG  
SNNTTIVLKLITLTLTPDSQOTGLSLVYSPTRANSTSEKKLPQPSKGVIAVA  
VIVCTVLAVIYGAALVFLYKKGKGLPCGRSGNQETTLPTPKRSEVVEBKSLKLEPENA  
LLQSSNGDKRAPRQGEKEYIDLRH"

BASE COUNT 708 a 735 c 783 g 654 t  
ORIGIN

Query Match	68.4%	Score 1333;	DB 11;	Length 2890;
-Best Local Similarity	80.7%;	Pred. No. 6.3e+280;		
-Matches 1571; Conservative	0;	Mismatches 370;	Indels	Gaps
OY 1 atggaggtcccaaggctgctgcgcctcttcgcgcgcgcgtgctgctgcctcgc				
Db 20 ATGGGGGTGCCCCAAACTGGTGTGCCTTTCTTGTCGCGCCTGCTGCTGCGCCGT				
OY 61 gtccgaggctgtgcccggaagagctlgaaagcctg-----cgccctgaagctgttgaagt				
Db 80 GCCCGGGGTGTCCAGAGAGAGGAAGAAGACAGCACATACCACGCCGCCGACTGGTGAAGCA				
OY 115 gaagtgggcagcacagcccctctcgaatgtcgycctctccacagtcaccaagcaactcagc				
Db 140 GAAGTGGGCAGACACAGCCCCTTCTCAAGTGTGGCCCCCTCACGGGCTCTCAGCAACTTCAGC				
OY 175 catgtcgcagctgtttctcgtccacaagaagaagcgacgtcatcttcgcgttgtgcag				
Db 200 CAAGTGCACCTGGTTTTTATTTCACAAAGAGAGAGCATCTGATTTTCCGTGTGCACCAA				
OY 225 ggcacagggccaagcgaaacctlgyggagtaagcagcagcgctcagccctccagacacaaggg				
Db 260 GGCAGAGGCCACACGGGGAACCTGTGAATATGAGCACCCCTTAGCCCTCAAGACTCGGTG				
OY 295 gctactctgcgcctgactccaagtcaccccccaagcagcgagcgacttctltlygcagggc				
Db 320 GCTACTCTGGCCCTTGAATCAGTCACTCTCCCTCCCTGATGAGCGAATGTTCTCTGTATAAGAGC				
OY 355 aagcgccctcagctcccaagagfacgcgcataccagctccgcgctctaaagaactccgaagag				
Db 380 AAGCACAACACGGCTCCAGGATACCTACGTTAGCTTTCAGGTTTCAAGCCCCAGAGAA				
OY 415 ccaaacatccaggccaacccccctgggcattccctltyaaacagbaagagocltgaagagtlc				
Db 440 CCMACTATTTCAAAGCCAATATGTCGTGGGCATTCCTATGTGACAGCAAGCAAGCTCAGGAGGTT				
OY 475 gctactctlgttagggagaagaoggttaaccocatctcttaagtcatactlgttaaagaatgyc				
Db 500 GCTACTCTGTGTGGAGAAACGGCTAACCCCATCTCTCAAGTCTATGTGTACAAAGAACGT				
OY 535 cgagctctgaagagaagaagaaccgggtlccaacatctcagctgcgtcccaagactlgaagtlc				
Db 560 CTGCCCTTGCAAGAGAGAGAGAACCGAGTTCTATTCACAGTCTATCATCAGATTGTCCAGTCC				
OY 595 agtltgtlgtlaaacctlvcagaagtalctltyaaagcacagctlgtltaaaagaacaagat				
Db 620 AGTGCCTTGTAACACTTGTAAGAGTGTCTGTGATGCACCCCTAGTTTAAGGAGACAAAGAT				
OY 655 gcccaagtttactlgtgagctcaacataccggcctgcgccagtlggaaccacatatgaagagtlc				
Db 680 GCCCAAGTTTAACTGTGAAGTCAAGCTAACCGGGCTACCCAGTGGGAACCATATGAAGAACTT				
OY 715 agggagaatcaaccgtccctcgtttctcaaccogaagaanaagtblgctlvgaaatlgaagccc				
Db 740 AAGGAGGTCTACCTGCTCCGTGTTTCTTACCTCGAAGAAAAGTGTGGGTGAGGTAAAGCTT				
OY 775 gltggaatgtclgaagaaagggaaccgctglgaaatcagtgltltygcltyatgcaacct				

D	800	GTGGGGCTGCTGAAGGAAGGAGATCATGTACATACTAGGTGTCTGCAGATGGCAACCT	855
Q	835	ccacacacatcagcatcgaagaagcagaaccacagcaccagagagagcagaagaaga	894
D	860	CAACCCCACTTCACTATCAACAAGAAGAGCCCAAGCACTGGGAGATGAGAAAGAGAGC	919
Q	895	accaagacaaacggagtcctgctgctgctgagacccctgcggcgaagaagaacagctggcgcat	954
D	920	ACCGATGAAATGGGCTCTCTCTTGTGAACTCTGGAAAGAGCAATGACCGGCTCTAC	979
Q	955	gaatlcagggccctgagacttgaacacacatactgctctgagtgaacacaagaacta	1014
D	980	CAGTGTCAAGTCTTGAGACGTGGAAACTACCATCACTGTCAAGTGAACCCCTGGAGCTG	1038
Q	1015	ctgtgtgaactatctgtctgaacgtccgaagtgaatcccgagccccctgaagaagaagaagc	1074
D	1040	CTGGGAACCTATGTGTCTGTGATGTTCAGTGAATTCACATGCCCCCTCAACTCCAGAAAGT	1099
Q	1075	agcaacccacaccctgagactctgtgaagcagaagtagcagaagactctgactcagctgctc	1134
D	1100	GAGACCTTCACGCTGACCTCGAGAGCGAAGTAACCAAGACTTGAATTTAGTGGCTG	1155
Q	1135	agaagaagacagggccaggtgtctggaagggggcctgtgcttcagttcgaatgaactgaa	1194
D	1160	AGAGCAAGAGACAGGCGACGCTGTGGAAAGCGTCCGCTCTCCAGCTAAMCAACGTGGA	1219
Q	1195	cggagagcagaagagcgcatcagctcgtcgctgctgtgtgccagacataccgagactgac	1254
D	1220	CGGAGAGCAGGGGAGCGGATCTCTGTGATGGCATCTGTCTCCAGAACTTCTTGCTTGAAT	1279
Q	1255	cgcacacagctgtgtcaacgtgtgccaattttgtgcccccttgatgtgcatcacaagagag	1314
D	1280	CGTACCCAGCTGTGTACAGCTGTGGGCAATTTTGGGCTCCCACTGATGTGATTAAGAGAGG	1338
Q	1315	aaggtgtgtgtgaagaagaaatggtgttgaaatctctgtgaagcgtcagaagcacc	1374
D	1340	ANGGTGTGGGTGCANAGAAATGCAGTCTGAATCTGTTGTGTAGGCTTCAGAGAACTCT	1399
Q	1375	cggccacacatctctctggaacgttcaacgagcagcagaatgaacaagaaccaagatccagag	1434
D	1400	CAGCCCACTTCTCTGGAATGTCAATGAGTTCCGGCAACTGAATGCAACCAATCACTCACAG	1458
Q	1435	cgagctcgtgaacaccctggaatgtctctgtgacccggagctgtgtgaagcaggttgtgaa	1499
D	1460	ACAGTAGTAGACACCTTGAATGTCTTGTGAGCCAGACCTTGTGGAGCAAGGTCAAGAG	1519
Q	1495	tgcagcgctccaacagcactgtggcacaacaccagacatacctctctctgtagagctgtcat	1554
D	1520	TGTACAGCTTCAACTCCTCGTGGGTTCAAAACCAACCAACCATTTGTTGAAGCTGGTCACT	1579
Q	1555	ttaacacacctccacacagactctcaacacaaaccaactgtgctcagcaatctcaactgtcagt	1614
D	1580	TTTAACCACTTCATACCTGATCTCAGGCAAAACCACTGGGCTCAAGCACTCAACTCACT	1638
Q	1615	ccctcatcagaagcacaagcaccctccacagagagaagactctgcagagcaggaagccgg	1674
D	1640	CCTCAACCAAGCCCAACAGACTCTCACAGAAAAAAGCTGCAACGCCAAGAGCAAA	1699
Q	1675	ggcgtgtgcatcgtgtgctgtgattgtgtgcatccctgtgtccctgtgcgtgtgcgtgtc	1734
D	1700	GGTGTGTATCTGTGCGGTGTGATAGTTGTATCACTTGTGCTGTCTGTGCTGGTGTCTGCT	1755
Q	1735	ctctatctctctctaagaagaaggaaactcgtgtcagagcgctcagaagagcagaagaatc	1799
D	1760	CTCTATTTTCTCTACAAAGAAAGGCAAACTCCCATGTGGAGGCTCGGGAATAAGAGATC	1819
Q	1795	acgctgcccccgctctgttaagcgaacctgtgtgtgaagttlaagtcagaatlaagtcaca	1854
D	1820	ACGCTGCCCGCACTCGTAAGAGTGAATTTGTAGTTGAAGTTAAGTCACATTAAGCTCCA	1879
Q	1855	gaagagatgggctctctgcagagcgacagcgtgtgacaaagaagctctcgggagacagga	1914
D	1880	GAAAGTAAAGGCTCTCTTACAGGCGACCAAGGTGAACAAGGGCTTCCAGAAACCAAGGA	1938



QY 1915 gagaatacatcgatctgagcattag 1941  
|||||  
Db 1940 GAGAAATACATCATGCTGCAGCATTAG 1966

RESULT 6  
AB035507 1988 bp mRNA ROD 03-OCT-2000  
LOCUS Rattus norvegicus mRNA for s-gicerin/MUC18, complete cds.  
DEFINITION AB035507  
ACCESSION AB035507.1 GI:10566950  
VERSION s-gicerin/MUC18.  
KEYWORDS Rattus norvegicus (strain:Sprague-Dawley) male heart cDNA to mRNA.  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 1988)  
AUTHORS Taira,E., Okumura,S. and Miki,N.  
TITLE rat s-gicerin/MUC18  
JOURNAL Published Only in Database (2000) In press  
REFERENCE 2 (bases 1 to 1988)  
AUTHORS Taira,E., Okumura,S. and Miki,N.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases. Elitai Taira, Osaka University Medical School, Department of Pharmacology,  
2-2 Yamada-oka, Suita, Osaka 565-0871, Japan  
(E-mail:elaita@pharm.med.osaka-u.ac.jp, Tel:+81-6-6879-3521,  
Fax:+81-6-6879-3521)

FEATURES  
source  
location/Qualifiers  
1..1988  
    /organism="Rattus norvegicus"  
    /strain="Sprague-Dawley"  
    /db\_xref="taxon:10116"  
    /sex="male"  
    /tissue\_type="heart"  
    17..1837  
        /gene="s-gicerin/MUC18"  
    17..1837  
        /codon\_start=1  
        /product="s-gicerin/MUC18"  
        /protein\_id="DABI6049.1"  
        /db\_xref="GI:10366951"  
        /translation="MGLEPLVCAFLFAACCRCRSRGVGEKOPPTLPDFVEYEVEN  
TALLCSGPAHSGNFSQEWELIHKERQIPFRVHQGGSEPEYEHRLSLHGPGAT  
LALCGVRPHDDRMFLCKSKOPRPQDHVVLOOVYKAPEPTIOAVNLGHVDIOELKEV  
ATCGANGVPRIPOVIWKNKPLOEEENRVIISSOTVESGLTIKSIVLSARVKED  
KDAPCFEIKRRSPSCNMKREKSVTPVLIPARKVWEYEPVLLKEGDHYKTRCLTE  
DGNPQFTTKRKNPSTEBMEBEESTDEGLISTEPAQHHSHGVTOCOSLDILETTVMIS  
SDPLELLVNYSVDVDPPTAPEVQEGDSLTITCKAESNOIDFELMDRTQLLGKGGR  
NILOSNNKREAGGRYLICVASVPSVGLNRTERRVSVGIFGSPMAAKREKVAAQENAMI  
ILSCGAAGHPQPTISMINGSATPMNDPPQTVVSTLVNLVPELLETGAECTASNSIGSA  
SYTYLIILKTFLTTLTHDSQTGTSTPYSPSRANSTSTKKLLQGEEKGVIVAA  
VIVETLVALVGLATLTYFYFKKRIPCBRSNGQENERTMT"

.BASE COUNT 503 a 544 c 553 g 388 t  
.ORIGIN

Query Match 61.8%; Score 1204.2; DB 11; Length 1988;  
Best Local Similarity 79.8%; Pred. No. 6.9e-25;  
Matches 1435; Conservative 0; Mismatches 356; Indels 6; Gaps 1;

QY 1 atggaggctcccaagctggtctgcgcttctgtcgcgcgcgctgtctgtctgtcctgcg 60  
|||  
Db 17 ATGGGCGTCGCCAGGCGCTGTGCGCCCTTCTATTCGCTGCCTGCTGCTGCAGC 76  
|||

QY 61 gtccgaggctgtgcccgaggagctgtgaag-----cttgccgctgaactgtgtgaagt 114  
|||  
Db 77 GCCACGGGTGTGCCAGGAGAAGAAAACAAGCCACACACCCAGCCGCGACCCGGTGAGGTA 136  
|||

QY 115 gaagtcgggcacacagcccttctaagatgtgagcctctcccagtlcccacaagcaacctcagc 174

Db	137	GAAAGTGGGCAACACGGCCCTTCTGAAGTGTGGCCCCACATCCCTTCAGGCAACTTTCACG	196
Qy	175	catgtcgcagctgatttctctgtccacaagaagaagcggacgcgtacatcttcctgtgtgcacag	234
Db	197	CAAGTGGAAATGSGTTTATTTCACAAAGSAGAGGCAGATACCATTTTCCGTGTGCACCCAG	256
Qy	235	ggccagggcccaagcgaacccctgtggagatcagcagcagcgtctccagcctcccaagacaagag	294
Db	257	GGCAAGGGCCAGAGTGAACCTGGGGAAATATGAACACCGCCTCAGCCCTCCAGGCCACGGA	316
Qy	295	gctactcttgcgccttgaccccaagtaaccccccaagaagcagcatcttctgtgccagggc	354
Db	317	GCTACTCTGTGGCCCTGTAGTCAAGTCACTCCCATATATACCGGATGTTCTGTGTAAAGAC	376
Qy	355	aagcgccctcggtcccaaggtacccagtcacagtcacagtcctcgctctacaagctccggagag	414
Db	377	AAGCAGCCACCGGCCCCAGATCCTACTATGTTAGCTGCAAGTCTACAAAGGCTCCAGAGGAG	436
Qy	415	ccaaacatccagtlcaacccccctgggcatccctgtgaacagtaagagccttgaagagtc	474
Db	437	CCAACTATCCAAAGCCAAAGCTGCTGGGCATCCACGTGAGACATACAGAGCTTAAGAGGTT	496
Qy	475	gctactctgttaggaggaagaaagggtaccaccatctctaaagtaactctgttacaagaatgagc	534
Db	497	GCTACTCTGTGTGGAGGAATAGGCTACCCCATTCCTCAATCTCATGTGTACAAGATGAGC	556
Qy	535	cggcctctgaaggaggaagaagaacccgggtccacatcagtcctgtcccaagcttggagatcg	594
Db	557	CGGCCCTTGCAAGAGAGGAGAAACCCGGTTCATATCCAGTGTGCACAGACCGTGGAGTCC	616
Qy	595	agltgtltgtcaaccttgcagaglatctctgaagcacaagcttgttlaaagaagacaagaat	654
Db	617	AGTGCTTGTAACACTTGAAAGTGTCTTAAGTGCACGGGTAGTCAAGSAGAACAAAGAT	676
Qy	655	gcccaagtttactctgtgagctaaactaccggctgcgcaggtgggaaccaatgaaagagttcc	714
Db	677	GCCCATGTTTACTGTGAAGTAGTACCGGGCCGCGAGTGGGAACCGCAAGAAGAAATCT	736
Qy	715	aagggaagtcacgcgtccctgcttcttaccgcgaacaagaanaagtgtggcttggaaagtggagccc	774
Db	737	AAGGAGGTCACGTGCCCTGTCTTATACCTGTGCAGAAAAAAGTGTGGGTGGAAGTATGAACCC	796
Qy	775	gttgggaatgtctgaagaaagggagccgcgttggaaatcagtglttlygcctatgtgcaacct	834
Db	797	GTGGGACGTGCTGAAGGAAGGGAGCCATAGTGAATCATGCTGTGATCGATCGGCCAACCT	856
Qy	835	ccaccaacactcagcatagagaacaaaccccagcacccggagagcagagaagaagaagaca	894
Db	857	CAACCCCACTTCACTATCAACAAGAAAGAACCCACGCACTGAGGAGATGGAAGAGGAGAC	916
Qy	895	accacaagcaaacgggttccgtgtgtctgtgagccttbcggagaagaacaacagttggcctat	954
Db	917	ACCGATGAAGAAGGTCCTCTCTCTGTGGAGCGGCCCCAGAAAGCACCAAGTGGGGTCTAC	976
Qy	955	gaatgtcagggccttggacttggacacccaatgatatcgtctgtgtgtgaaccaagaacta	1014
Db	977	CAATGTCAAGACCTTAGACTTGGAACATACAGTTATGCTGTGCGAGTGAACCCCTGTGAGCTG	1036
Qy	1015	ctgtgtgaacttgtgtctgaagtcgcaggttagtctccgcgaagcccttgaagacaagaaggc	1074
Db	1037	CTGTGGAACTATGTGTCTGATGTTTCAAGTGGATCAACTGCCCTGTGAAGTCCAGGAGGT	1096
Qy	1075	agcagctcacccttgacctgttgaagcagagtagccagaagccttgaattccagttgcgtgt	1134
Db	1097	GATATCCCTCAACACTGACTGTGAAGCGAGAGATTAACCAAGSACCTTAGTTGAGTGGCTG	1156
Qy	1135	agagaagaagacaagcccgaggtgtgtggaaaggggctgtgttcaagtcatgtaccctgaaa	1194
Db	1157	AGAGCAAGAAAGCGCAGCTTCTGTGGAAAGGGGCCCATCTTCAAGCTTAACAACAGTGAAG	1216
Qy	1195	cggagagcagaagcgcgtatcgtctgtgtgtgcgtcttctgtcccaagatacccgagcaaac	1254

Db 1217 AGGGAAGCAGGGGGAAGATATCTGTGTGTCATCTGTCCAGGCTTCTTGCTTGAAT 1276  
Qy 1255 GCGCAAGCCTGTGTAAGCTGTCATTTTGGCCCCCTTGTAGCTTCAAGAGAG 1314  
Db 1277 CCGACACGGCGGGGTACGCTGGGCAATTTTGGGCTCCCTGGATGGCAGCAAGAGAG 1336  
Qy 1315 AAGGTGTGTTGTAAGAGATATGTTGATGATCTGTCTGTGTAAGCTTCAAGAGAG 1374  
Db 1337 AAGGTGTGTTGTAAGAGATATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1396  
Qy 1375 GCGGCAAGCCTGTGTAAGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1434  
Db 1397 CAGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1456  
Qy 1435 GCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1494  
Db 1457 ACGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1516  
Qy 1495 TCGAGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1554  
Db 1517 TCGAGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1576  
Qy 1555 TCAAGCCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1614  
Db 1577 TCAAGCCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1636  
Qy 1615 CCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1674  
Db 1637 CCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1696  
Qy 1675 GCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1734  
Db 1697 GCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1756  
Qy 1735 CCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1793  
Db 1757 CCGGCTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 1815

RESULT 7  
AB035509 2772 bp mRNA ROD 03-Oct-2000  
LOCUS AB035509  
DEFINITION Mus musculus mRNA for s-glycerin/MUC18, complete cds.  
ACCESSION AB035509.1 GI:10566954  
VERSION s-glycerin/MUC18.  
KEYWORDS Mus musculus (strain: BALB/c) male cDNA to mRNA.  
SOURCE Mus musculus (strain: BALB/c) male cDNA to mRNA.  
ORGANISM Mus musculus (strain: BALB/c) male cDNA to mRNA.  
REFERENCE 1 (bases 1 to 2772)  
AUTHORS Taira, E., Okumura, S. and Miki, N.  
TITLE Published Only in Database (2000) In press  
JOURNAL 2 (bases 1 to 2772)  
AUTHORS Taira, E., Okumura, S. and Miki, N.  
REFERENCE Direct Submission  
AUTHORS Taira, Osaka University Medical School, Department of Pharmacology;  
TITLE Submitted (03-DEC-1999) to the DBJ/EMBL/GenBank databases. Elich  
JOURNAL 2-2 Yamada-Oka, Suita, Osaka 565-0871, Japan  
JOURNAL (E-mail: etaira@pharm.med.osaka-u.ac.jp, Tel.: +81-6-6879-3521,  
Fax: +81-6-6879-3521)  
FEATURES  
source 1. 2772  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="BALB/c"  
/db\_xref="taxon:10090"  
/sex="male"  
20. 1840  
/gene="s-glycerin/MUC18"  
20. 1840  
/gene="s-glycerin/MUC18"  
/codon\_start=1

/product="s-glycerin/MUC18"  
/protein\_id="BAB16051.1"  
/db\_xref="GI:10566954"  
/translation="MGLPKLVCFILPAACCCRRAGVPEKQKQPEVEHRLSLODSVAT  
TALCKPSRASNSQVDWFLHKRQLITRVHOGKQKQPEVEHRLSLODSVAT  
LALSHVTPDERMFLCKSRPLRQDHYVLOVFAKEEPTIOANVGHVQKLEVE  
ATCVGRNGYPIPOVLMVKNLSLROEENSVHIOSSQIVSSQIVLTKSVLSRLVED  
KDAOFYCELSYRPLSGNHMKESKEVYVPEYFAKEVWVEVEVGLKEDHYIRCLT  
DGNPOFHTNKRKDPSTGEMEESESTENLISLEPEKHSGSLYOCOSLDLETTLS  
SPLELVNVYSDVOYNPAPYQGESSTLTCFAESNDLRFEMLRDQTLCKGP  
VLOLVNVRERAGRYLCMASVPRVPLANTQVLSVGTESPMALKEKRVYQENAVL  
NLSGSGHPQPTISNVNVSATFEMNPQVTVSTLVNLTVELLETAEGCTASVLSG  
SNMTTIVLKVLTITLIPDSQVTSITFTVSPHNTSTSEKRLPOESKGVIVA  
VICTVLVLVGLALFLYKKGKLPCKGRSGKQEMERNIS1"  
BASE COUNT 675 a 706 c 749 g 642 t  
ORIGIN  
Query Match 61.4%; Score 1197.8; DB 11; Length 2772;  
Best Local Similarity 79.5%; Pred. No. 1.6e-250;  
Matches 1431; Conservative 0; Mismatches 362; Indels 6; Gaps 1;  
Qy 1 atggagcttccagagctgtgtctgtctgtctgtctgtctgtctgtctgtctgtctgtc 60  
Db 20 ATGGGGGTGCCCCAAACGAGGTGCTTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 79  
Qy 61 gtccgaggtgtgtcccgagagctgtgagcagctgt-----cgctgagctgtgtgaggtg 114  
Db 80 GCCCGGGGTGTGCCAGAGAGAGAAAGCAGCAGTACCAGCCAGCAGCTGTGTGAGGCA 139  
Qy 115 gaagtggcagacagacccctctgaagtgcgacctccagctccagctccagctccagctccag 174  
Db 140 GAATGGGACACACAGCCGCTTCAATGAGGCTTCAATGAGGCTTCAATGAGGCTTCAATGAG 199  
Qy 175 catgtgactgtgttctgtccaaagagagagagagagagagagagagagagagagagag 224  
Db 200 CAAGTGGAGCTGTGTTTGTGATTCACAAAGAGAGAGAGATGATGATGATGATGATGATGAT 259  
Qy 235 ggcagagggcagagcagacctgtgagagtaagcagcgtctcagcgtctcagcgtctcagagag 294  
Db 260 GGCAAGGGCCAGGGGAGAACCTGGTGAATATGACAGCCGCTTGGCTTCCAGAGAGAGGTT 319  
Qy 295 gctactctgtgacctgtactcaagtaagtaagtaagtaagtaagtaagtaagtaagtaag 354  
Db 320 GCTACTGTGGCCCTGAGTACAGTCACTCCCATGATGAGGAGATGTCCTGTCTAGAGAGC 379  
Qy 355 aagggccctgtgtccag 414  
Db 380 AAGGAGCCAGCGCTCCAGAGATCACTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGGAA 439  
Qy 415 ccaaacatccagagcagacccctgtgagcgtctgtgagagagagagagagagagagagag 474  
Db 440 CCAACTATTTCAAGCCATGTCGTGGCATCATGATGAGCAGAGGAGGCTTCAAGAGAGGTT 499  
Qy 475 gctactctgtgtgag 534  
Db 500 GCTACTGTGTGGGAGAGAACGGCTACCCCATCTCAAGCTCTTATGCTTCAAGAGAGAGT 559  
Qy 535 cggagctgtgag 594  
Db 560 CTGGCCCTTGCAGAGAGAGAGAACCGAGTTCAATTCAGATCATACAGATGTCGAGTCC 619  
Qy 595 agtgtgttgcacacctgtcagagatctctgaagcagcagctgtgttlaaagagagagagag 654  
Db 620 AGTGGCTTGTACACCTTGAAGAGTGTCTGAGTGCACAGCCTAGTTAAGAGAGAGAGATCT 679  
Qy 655 gccagatttactgtgagcctaagcagcagcagcagcagcagcagcagcagcagcagcagcag 714  
Db 680 GCCAGATTACTGTGAAGTACAGTACCGGCTACCGAGTGGAGAACCATGAGAGAGATCT 739  
Qy 715 agggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 774  
Db 740 AAGGAGGTCAGTCTGCTGTTTCTTACCTGCGAGAGAAAGTGTGGGTGAGGAGGCTT 799

[illegible][illegible]

QY 255 tgggagag taccgaacagcgtctcagctcccaagacagaggcgctcctctgagccctgagc 311  
 DB 485 AGATGAACCGGAGTACAGCAGAGCGGCTGTCGGGGGAGAGCAAGGCCCTGTGCTCATCAG 544  
 QY 312 tcaagtcaccccccagaaga-----gagcatctctctctgctcagggcaagcgccc 362  
 DB 545 CAAGGTGACGAGCAAGACAAAGCCAGAGACCTTCATCTCGACAGTTGGGGCGGACAGCCA 604  
 QY 363 tcggtcccaaggag taccgca tccagctccgcgtctacaagctcccgagagagccaaact 422  
 DB 605 GGGCGTGGGGGAGAGCGCCGACCGAGCTTACACCTCAAGATCCCGCACCCCTTAGAT 664  
 QY 423 ccaggttgaacccccctgggca tccctgtgaaacag taaaggacctggaagtgctacccg 482  
 DB 665 CACACCCAACTCGCGCGGCTCCCTGCAAAAGCAATGCATCTTAAAGATTGCTCAGTG 724  
 QY 483 tgaaggagaagacggtga cccca tctcctaagtcac tctgtaagaagatgagccgctct 542  
 DB 725 CACGACGAGAACAGCTTCCCGTCCCGCAACATCACCTGATCAAGAACGGGGAGCGCT 784  
 QY 543 gaaggaggagaagaacccggttcacac ttcagtcgtcccaagctgtagagtgagtggtt 602  
 DB 785 GCTGCAGGAGAGGACAGCAAGATCCTGACACGTTGGTCCGCGAGTCGAACGGGCT 844  
 QY 603 gtacaccttgcaagatctctgaaggacacagctggttaaagaagaacaaatgcccagtt 662  
 DB 845 GTACAGGTGTGTGAGCACTCTGTTACGAAGAGTGACACGGAGAGCGCAACTCCCTCTT 904  
 QY 663 ttaactctgagctaac taccgctgctccag tgggaacca tgaagagatccagggagat 722  
 DB 905 CCACCTGACCGCTCACACTACGCTGAGGACAGATGCCACCAAGACTCTCCGGGGT 964  
 QY 723 caccgtccctgcttttctaccggaacaaagtgctgagtgagtgagcc---cgtagg 779  
 DB 965 GAACGTCACGTGTTCTACCCACAGACGCTGAGAGCTGCGCGTGCCACCAACCGGG 1024  
 QY 780 aatgtcgaaggaagggacccgcg tggaaatcaagtg tctgctgta tggcaacctccac 839  
 DB 1025 GATAGTGAAGGAAGGGGATGATGTGAAGTTCGTCTGCAATGCGATGGGAATCCGGCGC 1084  
 QY 840 acaacttcagatc-----agcaagcagaacccacgacacccagggagcgagaaagagac 893  
 DB 1085 TGTCTTAGCTTTCAGGAGAGAGCTGGGGACACCTGCGAGATATGACTCTCGCTGCG 1144  
 QY 894 aaccaagcagaacgggtctcgtg tctgagcctgcccgaaggaaacacagtgaggcgctc 953  
 DB 1145 GATTAACCAACGATGAGTCTCTGATGCTGCAATGTCACCAAGCAGCAGCGGCTGTGA 1204  
 QY 954 tgaatgtcgaagggccttgagcttgagacac tgaatgcgtctgtagtgaacacagaaact 1013  
 DB 1205 CAGGTGCGACACACTGTGACTGTGACATATGACGACACAGGAAGGAGATGTGAGAGCTG 1264  
 QY 1014 actgtgtgaactatgtctgtagcgtccgagtgag tcccgagccctctgagagaacgaag 1073  
 DB 1265 GGTGACTATCATCTGAAGGGGTTCCAGTGAAGATGAGCCGCTCTCCATCATGAAG 1324  
 QY 1074 cagcagacctcaacctgaacctgtgagcgagagatgacagagacctgagatccagtggtc 1133  
 DB 1325 GGCACAGCTGAGGCTGAGCTGACCTGCCACACGCCCGTCAAGCTGAGATACCAATG--- 1381  
 QY 1134 gagaagaagacagacagcaggtgcttgaaagggggccttgctcagttcagatcagacagcga 1193  
 DB 1382 GAGGAGACGGAGGAGGAGAGAGTGGCAAGAGCAACAGACTCTCTCAGCAACACTGAC 1441  
 QY 1194 acggaaggcagaagcgagctac tgcgtcg tggcgctgtctgtgcccagacatacccgagtgaa 1253  
 DB 1442 CTTCGAGACTCCAGCAACTTCAAGCTGACAGAGTGAAGGCTCGCAGTGTCCGGGCTGGA 1501  
 QY 1254 cgcgaacagcgtgtgtaagtgagcatttggcccccttgatgagcatctaaagagag 1313  
 DB 1502 GCAGACCAACGACGAGTGGCGGTGCTGAAGGGGAAAGCCGCGCATCTGGCATCAGCCG 1561

QY 1314 gaagctgtgggtgaaagaatalgtgtgtgaatctgtcttgttaagcgttcaggcgacc 1373  
 DB 1562 CCGCGTGTACGTGCGCGGACAGCAAGATGATCACTGACCTGCAAGGCCATCGCCTTCCC 1621  
 QY 1374 ccggcccaacatctctctgtgaagcttcaacgacgagcgaagtgaacaagaacaaatccaca 1433  
 DB 1622 CCAGCCCTCTCTCCACTGAGACATCAACGGGAGC---ACCACACAGATCAATGAGAACCA 1678  
 QY 1434 ggcagctctgagcaccggaatgtctctgtgaccccgagctgtgtgagagagtgltga 1493  
 DB 1679 GCACATGGCCACACACCTGAGAGGTGCTGTGAGCCACACCTGCTGGGAGGAGGCAT 1738  
 QY 1494 atgcagggcctcccaacgacacccctgagcaaaaacacacagacatctctctctgagctgtcaa 1553  
 DB 1739 GTGCCGGGTGCACACGCGCTGGGGGTGAGGAGAGACATTCAGCGTGGT----- 1790  
 QY 1554 tttaacacccctacacacagactccacaacaacacatgctgctcagcacttccactgcag 1613  
 DB 1791 ----- 1790  
 QY 1614 tccctacacagagccaacagacactccacagagagaagctgcccagccgagagcg 1673  
 DB 1791 -----GATCAAAAGCCATTCGAGACAA 1813  
 QY 1674 gggcgtgtgacatcgtgtgctgtgtagtctgtgcatcctgtgtccctggcgggtgtgtgctgt 1733  
 DB 1814 AGGCATCATCATCGTGGGCATCATCATCTGCATCTGCTGGTGTGGCCGCTGGGCTCAT 1873  
 QY 1734 cctctatctctctctaaagaagggcaagctgcgtgtgacggcgtccagggagacagagat 1793  
 DB 1874 CATTTACTTCTGTGACAAAGAGGCGCAAGATCTCTGCGCGCGCGCAAAACGAGCAT 1933  
 QY 1794 cagcgtcccccgctgtgtaagaggaactgtgtgtagtgaagtcaagtaagctccc 1853  
 DB 1934 CACAAGCCCAAGAGCAGCTAAAGACAGATGTAGTTAACTTAAGTCAGATTAACATTTTC 1993  
 QY 1854 agaagaatgtggcctctctgcagggcagcagcgtgtgacagaagggctccggagagacagag 1913  
 DB 1994 CGAAGAGCGGGGCGCTCTGCAAGGGTGCACAGGGGAGAGAGAGTGGCGCGCTGACACAG 2053  
 QY 1914 agagaataacatcgatctgagagcatla 1940  
 DB 2054 CGAGAAATACATCGATCTGAGAAACTA 2080  
 RESULT 9  
 CHKLGP  
 LOCUS 2584 bp mRNA VRT 10-FEB-1999  
 DEFINITION Chicken mRNA for 1-glycerin precursor, complete cds.  
 ACCESSION D49849  
 VERSION D49849.1 GI:1008910  
 KEYWORDS 1-glycerin precursor; cell adhesion molecule; immunoglobulin superfamily.  
 SOURCE Gallus gallus (strain:Leghorn) adult gizzard cDNA to mRNA, clone\_lid:lambda gell.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 REFERENCE 1 (bases 1 to 2584)  
 AUTHORS Taira,E.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAR-1995) to the DDBJ/EMBL/GenBank databases. Elchi Taira, Osaka University School of Medicine, Pharmacology; 2-2 Yamadaoka, Suita, Osaka 565, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp, Tel:81-6-879-3521, Fax:81-6-879-3529)  
 REFERENCE 2 (bases 1 to 2584)  
 AUTHORS Taira,E., Nagino,T., Tanjura,H., Takaha,N., Kim,C.H., Kuo,C.H., Li,B.S., Higuchi,H. and Miki,N.  
 TITLE Expression and functional analysis of a novel isoform of glycerin, an immunoglobulin superfamily cell adhesion molecule  
 JOURNAL J. Biol. Chem. 270 (48), 28681-28687 (1995)  
 MEDLINE 96081930





QY	1027	gfgctcagctccagatgagctccgcagacccctgagaaacgaagaagcagaagcctcaac	1086
Db	928	gtgtctacactccgagatgtcccccagacccttgaaagacaaagagcgcctctacc	987
QY	1087	ctgacctgtgaggaagatgacccagacactgagctcagctggctgtagaagaag	1141
Db	988	ctgacctgtgaggaagatgacccagacactgagctcagctggctgtagaagaag	1047
QY	1142	-----	1141
Db	1048	cccgaggggctggggcaaggggctgaggggtgggtgggctggcgctgacaaagatg	1107
QY	1142	-----	1164
Db	1108	tgccttctccagacccttgaccacatctctatccacagaccagctgctggaagc	1167
QY	1165	gggcctgtgctcagctgcatgctcagacccgaaagcagagcagggcgcgctatgcctgcgctg	1224
Db	1168	gggcctgtgctcagctgcatgctcagacccgaaagcagagcagggcgcgctatgcctgcgctg	1227
QY	1225	ggcgtctgtgccagcataccgcgaactgaaacgcgaacagctgtgtcaacgtgtggccatttt	1284
Db	1228	ggcgtctgtgccagcataccgcgctcgaaccacacacacactggtcaagctggccattttt	1287
QY	1285	gg 1286	
Db	1288	gg 1289	

RESULT 11	AP001003/c	LOCUS	DEFINITION	AP001003	173038 bp	DNA	HTG	30-MAY-2000
			Home sapiens chromosome 11 clone RP11-680C5 map 11q23, WORKING DRAFT SEQUENCE, in unordered pieces.					
				AP001003				

KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Homo sapiens DNA, clone:RP11-680C5.
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 173038)

TITLE	Fujiyama, A., Yada, T., Tokochi, Y., Matsuda, H., Nishiguchi, E.,
JOURNAL	Homo sapiens 173,038 genomic DNA of 11q23
REFERENCE	Published Only in Database (2000) In press 2 (bases 1 to 173038)

TITLE  
JOURNAL

COMMENT

```

-----Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: haitori@gsc.riken.go.jp
-----Project Information
Center project name: HumDraff11
Center clone name: RP11-680C5
-----Summary Statistics
Sequencing vector: PCR products: 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 149676 bases at least Q40
Consensus quality: 160868 bases at least Q30
Consensus quality: 167161 bases at least Q20
Insert size: 170338; sum-of-configs

```

Quality coverage: 4.33x in 920 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

15559	15558	contig of	15558	bp	in	length	15558	bp	in	length
15659	33689	contig of	18031	bp	in	length	18031	bp	in	length
33790	47432	contig of	13643	bp	in	length	13643	bp	in	length
47393	63392	contig of	15860	bp	in	length	15860	bp	in	length
63493	74334	contig of	10742	bp	in	length	10742	bp	in	length
74335	85357	contig of	11023	bp	in	length	11023	bp	in	length
85458	93597	contig of	8140	bp	in	length	8140	bp	in	length
93698	101983	contig of	8286	bp	in	length	8286	bp	in	length
102088	110022	contig of	7939	bp	in	length	7939	bp	in	length
110123	118116	contig of	7994	bp	in	length	7994	bp	in	length
118117	126646	contig of	8248	bp	in	length	8248	bp	in	length
126655	132136	contig of	5572	bp	in	length	5572	bp	in	length
132337	135290	contig of	3054	bp	in	length	3054	bp	in	length
135391	140229	contig of	4839	bp	in	length	4839	bp	in	length
140330	143896	contig of	3567	bp	in	length	3567	bp	in	length
143997	147572	contig of	3576	bp	in	length	3576	bp	in	length
147673	150464	contig of	2774	bp	in	length	2774	bp	in	length
150547	153715	contig of	3169	bp	in	length	3169	bp	in	length
153816	157273	contig of	3458	bp	in	length	3458	bp	in	length
153734	159817	contig of	2444	bp	in	length	2444	bp	in	length
159918	161743	contig of	1826	bp	in	length	1826	bp	in	length
161844	163684	contig of	1841	bp	in	length	1841	bp	in	length
163785	165216	contig of	1432	bp	in	length	1432	bp	in	length
165317	166657	contig of	1341	bp	in	length	1341	bp	in	length
165758	168307	contig of	1550	bp	in	length	1550	bp	in	length
168408	170030	contig of	1623	bp	in	length	1623	bp	in	length
170131	171725	contig of	1595	bp	in	length	1595	bp	in	length
171826	173038	contig of	1213	bp	in	length	1213	bp	in	length

\* NOTE: This is a 'working draft' sequence.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

FEATURES	source
location/Qualifiers	location/Qualifiers
	1..173038
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="11"
	/clone="PP11-680C5"
	/map="11q23"
	1..15558
misc_feature	/note="assembly-fragment
	15659..33669
misc_feature	/note="assembly-fragment"
	33790..47432
misc_feature	/note="assembly-fragment"
	47533..63332
misc_feature	/note="assembly-fragment
	63493..74234
misc_feature	/note="assembly-fragment"
	74335..85337
misc_feature	/note="assembly-fragment"
	85458..93597
misc_feature	/note="assembly-fragment"
	93698..101983
misc_feature	/note="assembly-fragment"
	102084..110022
misc_feature	/note="assembly-fragment"
	110123..118116
misc_feature	/note="assembly-fragment"
	118217..126464
misc_feature	/note="assembly-fragment"
	126555..132136
misc_feature	/note="assembly-fragment"

```
misc_feature 132237..135290
/note="assembly_fragment"
misc_feature 135391..140229
/note="assembly_fragment"
misc_feature 140330..143896
/note="assembly_fragment"
misc_feature 143997..147572
/note="assembly_fragment"
misc_feature 147673..150446
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 150547..153715
/note="assembly_fragment"
misc_feature 153816..157273
/note="assembly_fragment"
misc_feature 157374..159817
/note="assembly_fragment"
misc_feature 159918..161743
/note="assembly_fragment"
misc_feature 161844..163684
/note="assembly_fragment"
misc_feature 163785..165216
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 165317..166657
/note="assembly_fragment"
misc_feature 166758..168307
/note="assembly_fragment"
misc_feature 168408..170030
/note="assembly_fragment"
misc_feature 170131..171725
/note="assembly_fragment"
misc_feature 171826..173038
/note="assembly_fragment"
BASE COUNT 45530 a 39606 c 41002 g 44200 t 2700 others
ORIGIN
```

```
Query Match 12.5% Score 243.2; DB 76; Length 173038;
Best Local Similarity 63.2% Pred. No. 4e-43;
Matches 722; Conservative 0; Mismatches 3; Indels 417; Gaps 6;
```

```
QY 560 ggggtccacatcagtcgcccagactgtagagtcgaagtggtttgtaacacctgacagagta 619
|||||
Db 33459 GGGTCCACATTCAGTGTGCTCCAGACTGTGAGTCGATGTTGTACACCTTGACAGAGTA 33400

QY 620 ttctgaaggacacagctggtttaaagaacaaagatgccagtttactgtgagctcaact 679
|||||
Db 33399 TTCTGAAGGACAGCTGTTAAAGAAAGACAAAGATGCCAGTTTACTGTGAGCTCACT 33340

QY 680 accgagctgccagtgaggaaacacatgaaggagtcacaggaagtcacagctccctgtttct 739
|||||
Db 33339 ACCGGGTGCCCACTGAGTGAAGACACATGAGAGAGTCAGGAGAGTCACTGCTGTTTCT 33280

QY 740 ----- 739

Db 33279 GTGAGTACTAACTGACTCTCTGCTTACTGGGGCTGGGGCTGCTGGAACCTCTGCTGGC 33220

QY 740 -----a 740
|
Db 33219 TCCCCCTTATCAGCCAGGTTGACTATGGGCCAAGAGTACCCCTTTGCCCTATGCACA 33160

QY 741 cccgcagaaaaaagtgtagctggaagtgagccgtggagatgctgaaggaaggagacgg 800
|||||
Db 33159 CCCGACAGAAAGTGCTGGCTGGAATGAGCCCGTGGGAATGCTGAAGAGGAGGAGCG 33100

QY 801 cgtggaaatcaggctgttggctgtagtggaacacctccacacacttagatataagcaagc- 859
|||||
Db 33099 CGTGGAAATACGCTGTTGGCTGATGGCAACCTCCACACACTTGAGATCAGCAAGCA 33040

QY 860 ----- 859

Db 33039 GGTGTGGGGCTTTTGCTGCTGGTGCAGGTGGGATGCACAAAGACCAAGTGAGTG 32980

QY 860 -----agaaccacagcaccaaggagga 882
```

```
|||||
Db 32979 CGGAGGTGCACTGTGCTCACTCCCAATCCCTCCCTAGAACCCCAAGACAGGAGGCA 32920

QY 883 gaggaaagacaacaaacagcaacaggggtccctggtgagacctcccgaaagaacac 942
|||||
Db 32919 GAGGAAGACAAACCAACGACGACGAGGGGTCTGCTGTGAGGCTCCCGGAAGGAACAC 32860

QY 943 agtggagctatgaatgtagagggcctgacttgaaacacatgatccgtcgtgagtga 1002
|||||
Db 32859 AGTGGGCTATGATTCAGGGCTTGACTTGAGACCATGATTCGCTGCTGATGAA 32800

QY 1003 ccaacgaactactgtgaaact----- 1024
|||||
Db 32799 CCACAGAGAACTACTGTGAACTGTGAGGGCTGGGACCCAGACAGAGGAGACAGGCTG 32740

QY 1025 -----atg 1027
|||||
Db 32739 GGGCAATAGGAACTGGCCCCACCCCTGCTGGGTGACATGATTGTTGCCCAAGATG 32680

QY 1028 tgtctgacgtccagtgagtcgccgcaagccctgagagacagaaagcaagcctcaccc 1087
|||||
Db 32679 TGCTGACGTCCGAGTGAATCCCGACAGCCCTGAGAGACAGAGGAGCAGCTCA-CC 32621

QY 1088 tgaacctggaagcagagagtagccaagacctgagttcagtggtcgaagaag----- 1141
|||||
Db 32620 TGACCTGTGA-GCAGAGAGTAGCCAGAGCTCGAGTCCAGAGCTGTGAGAAAGAGGTAC 32562

QY 1142 ----- 1141

Db 32561 CCAAGGGGTCTGGGCGCAAGGGCTGATGGAGGTGGGCTGGCGGTGACAAAGATG 32502

QY 1142 -----agaacagcagtgctgaagaag 1164
|||||
Db 32501 TGCTCTTCCAGACCCCTGACCATCTCTCATCCAGACAGGCCAGTGTGGAAGG 32442

QY 1165 gggcctgtgctcagttgcatgacatgaacggaagcaggaagcggctatgctgctg 1224
|||||
Db 32441 GGGCTGTGCTTCACTGTGCAATGACCTGAACGGGAGCAGAGGCGGCTATCCTGCTG 32382

QY 1225 ggcgtgtgcccagacatccccggagactgaaccgacacagctggtcagctggtcatttt 1284
|||||
Db 32381 GCGTGTGCCCCAGATACCCTGGGCTGACACGACAGCTGTCATCGTGCATTTTT 32322

QY 1285 gg 1286
||
Db 32321 CG 32320
```

```
RESULT 12
GHEMCAAC 1988 bp mRNA VRT 06-FEB-1997
LOCUS G.gallus mRNA for HEMCAM, short cytoplasmic form.
DEFINITION Y08856
ACCESSION Y08856
VERSION Y08856.1 GI:1621227
KEYWORDS chicken.
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1988)
AUTHORS Vainio,O., Dunon,D., Aissi,F., Dany,J.P., McNagny,K.M. and
Imhof,B.A.
HEMCAM, an adhesion molecule expressed by c-kit+ hemopoietic
progenitors
JOURNAL J. Cell Biol. 135 (6 Pt 1), 1655-1668 (1996)
MEDLINE 97133433
REFERENCE 2 (bases 1 to 1988)
AUTHORS Dunon,D.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1996) D. Dunon, Université Pierre et Marie Curie,
Unité CNRS 1135, Equipe Adhesion et Migration Cellulaire, 9 quai St
Bernard, 75005 Paris, FRANCE
COMMENT Related sequence: D38559.
```





KEYWORDS	adhesion molecule; HEMCAM, immunoglobulin supergene family.
SOURCE	chicken.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus. 1 (bases 1 to 2075)
AUTHORS	Vainio, O., Dunon, D., Aissi, F., Dangy, J.-P., McNagny, K.M. and Imhof, B.A.
TITLE	Imhof, B.A. HEMCAM, an adhesion molecule expressed by c-kit+ hemopoietic progenitors
JOURNAL	J. Cell Biol. 135 (6 Pt 1), 1655-1668 (1996)
MEDLINE	97133433
REFERENCE	2 (bases 1 to 2075)
AUTHORS	Dunon, D.
TITLE	Direct Submission
JOURNAL	Submitted (15-OCT-1996) D. Dunon, Université Pierre et Marie Curie, Unité CNRS 1135, Equipe Adhesion et Migration Cellulaire, 9 rue St Bernard, 75005 Paris, FRANCE
COMMENT	Related sequence: D38559.
FEATURES	Location/Qualifiers
SOURCE	1..2075
	/organism="Gallus gallus"
	/strain="H.B19ov+."
	/db_xref="taxon:9031"
	/dev_stage="E13"
	/tissue_type="thymus"
	201..305
	201..1715
	/note="soluble form"
	/codon_start=1
	/product="HEMCAM"
	/protein_id="CAA70079.1"
	/db_xref="GI:1621232"
	/db_xref="SPTRMBL:O98923"
	/translation="MAGGRAAPRGGMGGMGCCPFLLILCCGAAGRELYVMPAV LEVINGTARLECSFSTPGNAFTSIMFYVRRPLSGEVVRHHASGVRIDEYSE RLSGEKALSIKVTRODNARTICVGADSDQGESERTLEYTKLPAPETIPNSA GTPGSNDMLKIACTSENSEFPSPNITWNGEPLDIEBKTTLTLTVRESNLTYIV VSLFSKYVTREDRNSLPHCYHWLQOMRKPSRYNVVFYPTFEHLERATNAGI YKEBDVDKLVCODGNPAPIVSEFRRELGSWMDSLTADINDGLMLHNYSKSSGL VKRCOTLDLMDTHGEDEVELVANYIEGVYKMERSPLLHGDSVRLSCTAHPKLLDY OMRPARGRVAEGNOLLTNLTPEFTSNFSCRKARSVPGLDGSKQVAAYVKGRPRIV AISAPLVRRKDEVTINLTCKAIAPFQPSFHMSINSTVARGYLMALDMGAVRG"
	306..1712

mat\_peptide 440 a 592 c 681 g 362 t  
BASE COUNT  
ORIGIN

```

Query Match      11.9% Score 232.4; DB 4; Length 2075;
Best Local Similarity 49.3%; Pred. No. 1.7e-40;
Matches 890; Conservative 0; Mismatches 866; Indels 48; Gaps 9;
```

OY	18	ggtcgcgccttccttgctgcgcgcgcgtgctgtctgtcctcctcgctcgagggctgtgtcccg	77
Db	245	GGGGGGCGGTGGGGGCTGCTGCCCTTCCTGCTGCTCCTGCTGCCTGCTGCGTGCAACGCGG	304
OY	78	agaagccttagcacgcctgcgcgcgtgaagctgtgtgaagtgaagtggcaacacacaccttc	137
Db	305	CAGCGTGGAGGTCMAACATGCGCACGACGACTCTGGAAGTGAAGATCGGAGACACACGACGCT	364
OY	138	gaagtgcgcctcctccagctccc--aaagcaacctcagcatatgctgcagcttgtttctgt	194
Db	365	GAAGTCGACAGCTTCCATCCTCTGGGAATGCTCTCCCTACCTCCATCGATCGAGTGTCTTAGT	424
OY	195	ccacaaggagaaggaagcaccatcttcctgtgtgcgcagagccaagggccaagagcgaacc	254
Db	425	CAACCGCGCGCCATCCGATCGAGTGGTAGAGCTGCTCATACCCACCGCGGTTGGCAT	484
OY	255	tggaggatgacagcagcgcctcagactcccaaagaagggagcta---ctcgagccctgac	311
Db	485	AGATTGAACAACGATACACCAGCGCGCTCTCCGTTGGGGGAGACAAAGGCGCTGTCCATTCAG	544
OY	312	tcaagtcaccccccaagaagca-----gcgatcttctgtgcaagggcaagcgccc	362

Db	545	CAAGGTGACCAAGGCAAGCAACGCCAGACCTTTCATCTCCAGGTTGGGGCGGACGCCA	604
Oy	363	tcggctcccaagtagtaccgcatccagctcccggtctataaagctccggagagcaaacgt	422
Db	605	GGGCGTGGGGGAGAGCCACCACCGACTCTTACCTTACAAATATCCCGCACCCCTTAGAT	664
Oy	423	ccaggtcaaccgccctgggcatcccttgtgaacagtaaggaagctgagagtgctacctg	482
Db	665	CACACCCAAATCTGGCGGGCATCTCTGCACAAAGAAATGACATTTGAAGATTCTCAGTG	724
Oy	483	tgtagggagaaacggygtaccatctctcaatgtacltgytataagaatgycgcgctct	542
Db	725	CACGAGCAGAAACAGCTTCCGTCGCCCAATCATCTGCTTACAAAGGGGAGCGCT	784
Oy	543	gaaggaagaagaacccggtgccacatctagtggtcccaagcagtgagtgtagtggtt	602
Db	785	GCTGCAGAGAGAGGACAAAGCAAGCAAAAGATCTGACCACTGTGTCCGGCAGTCGAACGGCT	844
Oy	603	gtacacctgtcagatattctcgaagcacagctggtttaaagaagacaagaatgcccagtt	662
Db	845	GTACACGGTGGTGAAGCACTGTTACACAAAGTGTACACGGGAGACCGCACTCCTT	904
Oy	663	ttactgtagctcaactacccggtctgccagtgaggaaacacatgaagagatccaggyaagt	722
Db	905	CCACAGCACCCCTGCACACTGCTGCAGGAGACAGATCGCACCAAGGACTCTCCGGGCT	964
Oy	723	caacgtcccttcttcttaaccgaagaanaaagtgctgtggaagtgaac---cgtggt	779
Db	965	GAACCTTCACGTGTGTTCACCCACAGAGACAGTGGAGCTCGCGTGGCAACACCCGG	1022
Oy	780	aatgtctgaaggaagggagcccgctggaatacagtgcttggtcgtatgtaaacccctccac	839
Db	1025	GATAGTGAAGGAAGGGATGATGTGAAGTTGGTCTGGATGCTGATGGGAATCCGGCGC	1082
Oy	840	acaattcagcatc-----agcaaggaaccaccagcaaccaagggagcagaggaagagac	893
Db	1085	TGTCCTTTAGCTTTCTTCAGAGAGAGAGCTGGGGGACACGTCGCAAGATATGACCTTCGCTGAC	1144
Oy	894	aaccaagaacaacggggtctctgtgtgtgtgagccctggccggaaggaacacagtggtccta	953
Db	1145	GGATATCCACACATGAGAGTCTCGATGCTGCACAAATGTACGACAGACGACGAGCGTCTGTA	1204
Oy	954	tgaatgtcaaggccctgtgacttgcacacaaagatacgtctgtatgtgaaccacaagaact	1011
Db	1205	CAGGTCCAGACACTGGACTTTGGACGATATGACGACGACGAAGGGGATGTGGAGCTGCT	1266
Oy	1014	actgtgtgaactatgtgtctgtacgtctcgagtgaatcccgacagcccttgagagaacagaag	1071
Db	1265	GGTGAACATCAATCCAAAGGGGTCCAGGTGAAGATGAGAGCGTCTCTCACCCCTTCATGAAGG	1324
Oy	1074	cagagacctccctctgagcctgtgtagagagaagatagcaagagaccttgatctcagtgct	1133
Db	1325	GGACAGCTGTGGCTGTGACTGTACATGCCACACCCCGTCAAGCTGGACTTACCAATG---	1381
Oy	1134	gagagaagaagacaagcagagtgctgtctgaagaagggccctgtgcttcagtgatgcactgaa	1193
Db	1382	GAGGGACCGCAGGGGCGAGAAAGTGTGGAGAAAGGCAACACAGCTCTCTCTACCAACTTCAC	1444
Oy	1194	acggagagcagagagcggtacatcgtgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	1253
Db	1442	CTTGAGAGCCTCCAGCAACTTTCAGCTGACAGATGAGAGGCTCGACAGTGTCCGGGGCTGGA	1501
Oy	1254	ccgcaacagctgtgtcaagtggtgcatcttctgcccccttgagatggtcatcaagagag	1313
Db	1502	GCAAGAGCAAGAGGTGGCCCTGTGCTGTGAAGGGGAAGCCGGCATTGTGGCCATCAGGCG	1561
Oy	1314	gaaggtgttggtgtagaagagatagtgtgtgaatctgtgttgaaagcgttaagggcacc	1373
Db	1562	CCGCTGTAGCTGGCGCAGGACCAAGTGTATCACTGACCTTCAAAGGCCATCGCTTCC	1621
Oy	1374	ccggcccaacatctccctggaagtgtaacggcaacggaaatgaaacaagaccgaatcacaa	1433

Db	1622	CCAGCCCTCTCTCCACTGGAGCATCAGCGGAGGGAAGG---CTGGGCGGGGTGCTGTG	1678
Qy	1434	gcgagtcctgagcaccctgaatgctcgtgaccccgagcgtgtgagacagtggtga	1493
Db	1679	GCGGCTCTCGGGATGGGGGCTGTGTGGCTGTGAGGCCCGGCTTTCGGTGTCCCT	1738
Qy	1494	atgcacggcctccacagacctgtggaanaaacaccacatccctctcctcgtgagctgtca	1553
Db	1739	ACAGACCACGAGTACATGAGAACACGACATGGCCACACCTGACGCTGCTGTGAG	1798
Qy	1554	tttaacacccctcacacacagactccaacaacacactgtgctcagactccaactgtccag	1613
Db	1799	CCACACACCTGTGCGGGGAGAGCCA-----TGTGGCGGGGTGCCAACGGCGTG	1847
Qy	1614	tccctataccagacgaacaaagactccaacagagaagaagcgtccggagccggagagccg	1673
Db	1848	GGGTCACGAGACACATCCAGCTGTGATCAAAAGCCA-----TCGGAGAGCA	1900
Qy	1674	ggcgctgtgcatcgtcgt	1733
Db	1901	AGGCATCATCATGTCGGCCATCATGTCATCTGCTGCTGTGTGTGTGTGTGTGTGTGT	1960
Qy	1734	cctctattcctctataagaaggaagcgtcgcgtgcaagcgtcaggaagcaagagat	1793
Db	1961	CATTACTTCTGCACAAAGAGGCAAGATCTCTGCGCGCCGACGCAACAGGACAT	2020
Qy	1794	cacg 1797	
Db	2021	AGCG 2024	

RESULT	14
CHKS	CHKSGP
LOCUS	2365 bp mRNA VRT 08-FEB-1999
DEFINITION	Chicken mRNA for s-gicerin precursor, complete cds.
ACCESSION	D38559
VERSION	D38559.1 GI:1009246
KEYWORDS	s-gicerin precursor; cell adhesion molecule; extracellular matrix protein; neurite outgrowth factor-binding; laminin family; immunoglobulin superfamily; integral membrane glycoprotein.
SOURCE	Gallus gallus (strain Leg Horn) adult gizzard (library: Lambda gII) cDNA to mRNA.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 2365)
TITLE	Taira,E.
JOURNAL	Direct Submission Submitted (20-OCT-1994) to the DBJ/EMBL/GenBank databases. Elitchi Taira, Osaka University School of Medicine, Pharmacology; 2-2 Yamadaoka, Suita, Osaka 565, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp, Tel:81-6-879-3521, Fax:81-6-879-3529)
REFERENCE	2 (bases 1 to 2365)
AUTHORS	Taira,E., Takaha,N., Taniura,H., Kim,C.H. and Miki,N.
TITLE	Molecular cloning and functional expression of gicerin, a novel cell adhesion molecule that binds to neurite outgrowth factor
JOURNAL	Neuron 12 (4), 861-872 (1994)
MEDLINE	94213753
REFERENCE	3 (sites)
AUTHORS	Taira,E., Nagino,T., Taniura,H., Takaha,N., Kim,C.H., Kuo,C.H., Li,B.S., Higuchi,H. and Miki,N.
TITLE	Expression and functional analysis of a novel isoform of gicerin, an immunoglobulin superfamily cell adhesion molecule
JOURNAL	J. Biol. Chem. 270 (48), 28681-28687 (1995)
MEDLINE	96081930
COMMENT	On Oct 6, 1995 this sequence version replaced gi:559700.
FEATURES	Sequence updated (05-Oct-1995) by:Elitchi Taira. Location/Qualifiers 1..2365 /organism="Gallus gallus" /strain="Leghorn" /db_xref="taxon:9031"

Db	1622	CCAGCCCTCTCTCCACTGGAGCATCAGCGGAGGGAAGG---CTGGGCGGGGTGCTGTG	1678
Qy	1434	gcgagtcctgagcaccctgaatgctcgtgaccccgagcgtgtgagacagtggtga	1493
Db	1679	GCGGCTCTCGGGATGGGGGCTGTGTGGCTGTGAGGCCCGGCTTTCGGTGTCCCT	1738
Qy	1494	atgcacggcctccacagacctgtggaanaaacaccacatccctctcctcgtgagctgtca	1553
Db	1739	ACAGACCACGAGTACATGAGAACACGACATGGCCACACCTGACGCTGCTGTGAG	1798
Qy	1554	tttaacacccctcacacacagactccaacaacacactgtgctcagactccaactgtccag	1613
Db	1799	CCACACACCTGTGCGGGGAGAGCCA-----TGTGGCGGGGTGCCAACGGCGTG	1847
Qy	1614	tccctataccagacgaacaaagactccaacagagaagaagcgtccggagccggagagccg	1673
Db	1848	GGGTCACGAGACACATCCAGCTGTGATCAAAAGCCA-----TCGGAGAGCA	1900
Qy	1674	ggcgctgtgcatcgtcgt	1733
Db	1901	AGGCATCATCATGTCGGCCATCATGTCATCTGCTGCTGTGTGTGTGTGTGTGTGTGT	1960
Qy	1734	cctctattcctctataagaaggaagcgtcgcgtgcaagcgtcaggaagcaagagat	1793
Db	1961	CATTACTTCTGCACAAAGAGGCAAGATCTCTGCGCGCCGACGCAACAGGACAT	2020
Qy	1794	cacg 1797	
Db	2021	AGCG 2024	

Query Match	11.9%	Score 232;	DB 4;	Length 2365;
Best Local Similarity	50.5%	Pred. No. 2e-40;		
Matches	760;	Conservative	0;	Mismatches 715; Indels 29; Gaps 7;

Qy	37	gcccgt	96
Db	265	GCCCCCTTCCTGTCCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	324
Qy	97	cctgagctgt	156
Db	325	CCAGCAGT	384
Qy	157	tc---ccaaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	213
Db	385	CTGGGAATGCT	444
Qy	214	ctcat-----cttcgt	268
Db	445	CGGGTAGGGGT	504
Qy	269	agcgt	328
Db	505	AGCGGCTGT	564
Qy	329	acgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	382
Db	565	ACGCCAGACCTTCT	624
Qy	383	tcacgtcgt	442
Db	625	CCGAGCTTACACCTTACACCTTACACCTTACACCTTACACCTTACACCTTACACCTT	684
Qy	443	tccctgt	502
Db	685	TCCCTGCACAAACCAATGACATTTTGAAGATTCTGAGTCCAGCAGCAGCAAGAGCTTC	744
Qy	503	cccttcctcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	562
Db	745	CGTCCCAACATCT	804
Qy	563	tcacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	622
Db	805	CANAAGATCTGACACGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	864
Qy	623	tgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag	682
Db	865	TGTTACAGAGGTGACACGGAGACCGCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCT	924

OY	683	ggctgcgcacgctgaggaaaccacatgaaagaggtccaaaggaagtaaccgctccctgcttctctcc	742
Db	925	ggctgcgcacgacagatgccaccacaaaggactctcccgggctgaacctcattgtttctaac	984
OY	743	cgcacgaagaaagatgttgcctcgaagatggagcc---cgtggaaatgcctgaagaaagagacc	799
Db	985	ccacagacagacagctggagctggccgctggggaccacacggccggggaattagtgaagaagggatg	1044
OY	800	ggctggaaatcgaagtgttttgctgatgtgcaaccctccacacacttcagatc-----a	853
Db	1045	atgtgaagtgtgctgtccgatgctgtgatgggaatccggccctctcttattgacttttcacaga	1104
OY	854	gcaagaagaagaccacagacacaggagagagagaaagajaaacaacaaagacaacaggtgtcc	913
Db	1105	gaagactgggggacacagctggcaggtatgactcccttgccgtatccaaacagatggagctcc	1164
OY	914	tggctgcctgagccttgcgccgaaagagacacagttggcgctatgatgttaagagccctgagct	973
Db	1165	tcgatgttcacaaatgtaacagcaagacagacagcggctgtgacagctggccacagacactggact	1224
OY	974	tggacacactgatatcgcgtctgagttgaaaccaagaaactactgtgtgaactatgtgtctg	1033
Db	1225	tggacagatattgacgacagcgcgaaggagattgtgacctggtgtgaaactatcattcgaagggg	1284
OY	1034	acgtccagagttagtcctccgagagccctctgagagacaaagaaagagcagcctctcaaccagact	1093
Db	1285	tccagctgaanagtatggacccgtctcttaacccctttatgaagaaggagacgctgtagcctgact	1344
OY	1094	gtcaagacagagagtagcaagaagacctcgaattccagtggcctgagaaagaaagacagccag	1153
Db	1345	gcactggccacagcccccgcgaactgtgacttaccatg---gaggagccgagaggggacaga	1401
OY	1154	tgtctgaaagaggggccttgctcgaattgtcatgacctgaacagggagagcagagagcgct	1213
Db	1402	acgtggcgaamaagcaaacagctctcttgcacacacacttgcagagacttccagacact	1461
OY	1214	atcgcctgcgtgcgtcctctgcccagacatacccgagactgaaccgcaacagacttgtaag	1273
Db	1462	tcacgttcgacagatgaaggctgcgaatgtgtccggccttgagacagacagacagatggcgcg	1521
OY	1274	tggccacttttggcccccttggatctgacatcaaaagaaagaaaggtgtgtgggtgaaga	1333
Db	1522	tggcgtgaaggagggaagccgcgcacatcgtggccatcagcagcccccgtgtgactggcgagg	1581
OY	1334	atatagttttgaatctgtcttgttgaaggtcaaggcaaccccgcccaactctctctgga	1393
Db	1582	acgaagatcatcaactgacctgcgaaggccatncccttcccacaccccttcttccacttga	1641
OY	1394	acgtcaagcgcgcgcgaagtgaacagacagaactccacagcagagctcctgagcccttga	1453
Db	1642	ggcgtacacggcagc---accacacagatcattatggaacacacacacatgcccacgcaactga	1698
OY	1454	atgtccctcgtgaccccgagcgttgttgaagacaagttgttaattgacaggtccctccaaagcc	1513
Db	1699	cgggtcctgctgagccacacactgctgctggcgacagagccatgtgcccgggtgtccaatggcc	1758
OY	1514	tggg 1517	
Db	1759	tggg 1762	
RESULT 15			
LOCUS	AP001557		
DEFINITION	AP001557 177595 bp DNA	HTG	30-MAY-2000
ACCESSION	AP001557		
VERSION	AP001557.2 GI:8117391		
KEYWORDS	HTG: HTGS. PHASE1: HTGS. DRAFT		
SOURCE	Homo sapiens DNA, clone:RP11-680A7.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

```

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 177595)
TITLE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
JOURNAL Fujiyama,A., Yada,T., Tokioki,Y., Watanabe,H. and Sakaki,Y.
REFERENCE Homo sapiens 177,595 genomic DNA of 11q23
AUTHORS Published Only in Database (2000) In press
TITLE 2 (bases 1 to 177595)
JOURNAL Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
REFERENCE Fujiyama,A., Yada,T., Tokioki,Y., Watanabe,H. and Sakaki,Y.
COMMENT Direct Submission
Submitted (20-MAR-2000) to the DDBJ/EMBL/GenBank databases.
Masahita Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.riken.go.jp, URL:http://hgp-gsc.riken.go.jp/,
Tel:81-42-778-9923. Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced g1:7380892.

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp-gsc.riken.go.jp/
Contact: hattori@gs.riken.go.jp

----- Project Information
Center project name: HumDrat11
Center clone name: RP11-680A7

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator PCR products; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 152340 bases at least Q40
Consensus quality: 165308 bases at least Q30
Consensus quality: 171688 bases at least Q20
Insert size: 174695; sum-of-contigs
Quality coverage: 4.19x in Q20 bases; sum-of-contigs

-----
NOTE: This is a 'working draft' sequence. It currently consists of
30 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 28368 contig of 28368 bp in length
28469 50818 contig of 22350 bp in length
50919 61024 contig of 10106 bp in length
61125 70311 contig of 9187 bp in length
70412 79377 contig of 8966 bp in length
79478 88609 contig of 9132 bp in length
88710 94861 contig of 6152 bp in length
94962 103809 contig of 8848 bp in length
103910 110603 contig of 6694 bp in length
110704 117602 contig of 6699 bp in length
117703 123711 contig of 6199 bp in length
123812 130010 contig of 6199 bp in length
130111 134042 contig of 3932 bp in length
134143 138065 contig of 3932 bp in length
138166 141127 contig of 2862 bp in length
141228 146317 contig of 2926 bp in length
146418 150691 contig of 5090 bp in length
150792 154193 contig of 4274 bp in length
154294 157006 contig of 3402 bp in length
157107 159710 contig of 2713 bp in length
159811 162554 contig of 2604 bp in length
162655 164588 contig of 2744 bp in length
164689 166160 contig of 1934 bp in length
166261 167844 contig of 1472 bp in length
167945 169717 contig of 1584 bp in length
169818 171073 contig of 1773 bp in length
171174 173341 contig of 1256 bp in length
173342 174526 contig of 2068 bp in length
174627 176098 contig of 1185 bp in length
176199 177595 contig of 1472 bp in length
1397 bp in length

Sequence updated (26-May-2000).

```



















```

FT XX /*tag= C
FN XX
AU 068723-B.
PD XX
19-FEB-1998.
XX
PF 02-DEC-1997; 97AU-0046867.
XX
PR 02-DEC-1997; 97AU-0046867.
XX
PA (KIRI ) KIRIN AMGEN INC.
XX
LIn F:
XX
DR WPI: 1998-261957/24.
DR P-PSDB: W58399.
XX
FT Recombinant human erythropoietin - potentially useful for diagnosis
PT and treatment of blood disorders
XX
XX
PS Example 3; Page 34-36; 100pp; English.
XX
CC The present sequence encodes monkey erythropoietin (EPO), from an
CC example of the present invention. The present invention describes
CC recombinant human EPO which causes bone marrow cells to increase
CC production of reticulocytes or red blood cells, where the polypeptide
CC is the product of expression in CHO (Chinese hamster ovary) cells of
CC an exogenous DNA sequence encoding human EPO. EPO is potentially
CC useful in the diagnosis and treatment of blood disorders
CC characterised by low or defective red blood cell production.
XX
SQ Sequence 1344 BP; 280 A; 408 C; 400 G; 256 T; 0 other;

Query Match 2.3%; Score 44.6; DB 19; Length 1344;
Best Local Similarity 50.7%; Pred. No. 0.078;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0.

QY 1059 tgaagacagaaaggaagcagacagctacacctgacctgtgagagagatagccaggaacct 1118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 617 TGATGCTTCGGAGTGGAGCAGCGCCGCGCATCTGGGAGGAGATGCGTTCCTGGGCTC 558

QY 1119 cgagttccagtgagtcgagagagaagacagggccaggtgtctggaagaaggccctgtctta 1178
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 557 CGAGGCGCCGGAAGCAGAGTGTGATGCTGCCAAGGCCACCTGATGCTTTATCCATGTGCA 498

QY 1179 gttgatcatcctgaaacgggaagcaggaaggcgctatcgctcggtggcgtgtgtgccag 1238
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 497 GCTGCAAGGGGCTCGAAGAGCTGGGAAGATTGGCCACACAGGCGCTGGCCCCGAGAGCAG 438

QY 1239 cataccggactgaaacgcagacacagctgctc 1269
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 437 CTTCTGAGAGCAGGGGCCAGGCCCTCCAGAC 407

RESULT 7
V21685/C
-ID V21685 standard; DNA; 5107 BP.
XX
XX V21685;
XX
DE 17-AUG-1998 (first entry)
XX
DE Vector plasmid CMV-KM-cMEPO.
XX
XX Polynucleotide delivery; plasmid CMV-KM-cMEPO; vector;
KW gene therapy; vaccine; polycationic agent; cynomolgus monkey;
KW erythropoietin; EPO; anaemia; ss.
XX
XX Chimeric - Macaca cynomolgus.
OS Chimeric - Human cytomagalovirus.
DS Chimeric - Bos taurus.
XX

```

PN	XX	W09806437-A2.
PD	XX	19-FEB-1998.
PE	XX	13-AUG-1997; 97WO-US14465.
PR	XX	13-AUG-1996; 96US-0023867.
XX	XX	(CHIR ) CHIRON CORP.
PA	XX	Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA;
PT	XX	Murphy JE, Tetsuo U, Zukermann R;
PT	XX	WPI; 1998-159296/14.
XX	XX	
PS	XX	Example 7; Page 81-83; 100pp; English.
XX	XX	
CC	CC	This polynucleotide sequence comprises plasmid CMV-KM-cmeEPO, a
CC	CC	vector in which cynomolgus monkey erythropoietin (EPO) cDNA
CC	CC	(ATCC 67545) is inserted into the XbaI and BamHI sites of the
CC	CC	mammalian expression vector CMVkm2 (see V21684) under control of
CC	CC	the human cytomegalovirus immediate early gene promoter/enhancer and
CC	CC	the polyA site of the bovine growth hormone gene. Hematocrit
CC	CC	levels increased in SCID mice injected with the vector 1.m.
CC	CC	pcMV-KM-cmeEPO is an example of a vector that can be used in novel
CC	CC	compositions and methods for improved polynucleotide delivery into
CC	CC	cells. In these methods, polycationic agents are used to increase
CC	CC	the frequency of uptake of a nucleic acid (see also V21683-86) into
CC	CC	a cell. The polycationic agent can condense with the nucleic acid
CC	CC	and inhibit serum and/or nuclease degradation of the nucleic acid.
CC	CC	The nucleic acid can be a vector, may express a therapeutic protein
CC	CC	or a vaccinating viral or cancer antigen, or is itself therapeutic
CC	CC	(antisenase or ribozyme). The methods and compositions can be used
CC	CC	in the gene therapy of many diseases.
SO	SO	Sequence 5107 BP; 1145 A; 1463 C; 1335 G; 1164 T; 0 other;
Query Match	2.3%; Score 44.6; DB 19; Length 5107;	
Best Local Similarity	50.7%; Pred. No. 0.13;	
Matches 107; Conservative	0; Mismatches 104; Indels 0; Gaps 0;	
QY	1059	tgaagacaaggaagcaagcagcctacccctgacacctgtgaaggcaagaaagtaagccaggacct 1118
DB	2187	TGATGCTTGGGAGTGTGACACGCGGAGCGCGCTCTGGGAGGAGATGCTTCTCGGCTC 2128
QY	1119	cgaagtcagtgagctgcaagaagaagaacagcgccagctgtgtgaaaagggagcctgtgtctca 1178
DB	2127	CCAGCGCCCGAAGACAGATGTGTGATCTGTGCAAGGCCACTGATGGCTTTATCCATGTCGA 2068
QY	1179	gttcgataagccttgaacagggaagcaggaagggcggtatcgtctgtgctgtgtgtgtccag 1238
DB	2067	GCTGAGGGGGCTCGAAAGCTGTGGAAAGTGTGGCCAAACAGCGCTGTGGCCGACGACAG 2008
QY	1239	cataccggactgaacgcgcacacagctgtgc 1269
DB	2007	CTTCTGAGAGCAGAGGCCACGCGCTCTCCAGAC 1977
RESULT	8	
ID	A10594	
XX	A10594 standard; DNA; 10732 BP.	
AC	A10594;	
DT	29-JUN-2000 (first entry)	
XX	Gene encoding a subunit of cellulose synthase.	

XX Cellulose synthase; cellulose production; increase yield; ds.  
 XX Vigna angularis.  
 OS JP2000060568-A.  
 PN 29-FEB-2000.  
 PD 26-AUG-1998; 98JP-0239998.  
 XX 26-AUG-1998; 98JP-0239998.  
 PR 26-AUG-1998; 98JP-0239998.  
 XX (MIZU/ ) MIZUNO K.  
 PA (OJIP ) OJI PAPER CO.  
 XX WPI; 2000-342371/30.  
 DR P-PSDB; Y85179.  
 XX  
 PT A gene encoding a cellulose synthetic equipment - for the improvement  
 PT in the amount of cellulose synthesised in a plant body  
 PS Claim 2; Page 14-21; 32pp; Japanese.  
 XX  
 CC This sequence represents a gene encoding a subunit of the cellulose  
 CC synthase complex of Vigna angularis. The invention relates to subunits of  
 CC cellulose synthetic equipment, that can be used to increase the amount of  
 CC cellulose synthesised by a plant. The proteins and genes encoding them  
 CC can also be used to improve the properties of the cellulose being  
 CC produced by a plant.  
 XX  
 SQ Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;

Query Match 2.2%; Score 42.8; DB 21; Length 10732;  
 Best Local Similarity 14.2%; Pred. No. 0.49;  
 Matches 113; Conservative 305; Mismatches 369; Indels 7; Gaps 2;

QY 427 gtaaccccttgaggcctccctgtgaacagaaagagcctgagagctcctcctgtca 486  
 DB 9162 snsrhaasnystrhaaargaaagaaagaaargagntcrgnaaagaaagaaagaaag 9221  
 QY 487 gggagaaagcgggtaccatcctcctcaatcctcgttacaagaatgacggcctcgaag 546  
 DB 9222 smrtlysasasnthrthlyrvaarvagysggargaaagaaagaaagaaagaaagaa 9281  
 QY 547 gaggagaagaacgggtcccaatcagtcgtcccaacagctgagtcgaagtcgtgtc 606  
 DB 9282 aagvaarhaysngngsrthrysaaysggcysnaargaryaaargyaa 9341  
 QY 607 acctgcagagatctctgaagcagacagctggttaagaagaaagaaagaaagaaagaa 666  
 DB 9342 aaysmlysaasngvamtlyragsrthrgargarsystrthaastrvagsrstraast 9401  
 QY 667 tgtgagctcaactacggcgtccagctgggaacacatgaagaaagtcagggaaagtcacc 726  
 DB 9402 hrartrgthrrthrysaasnthrthrgnycyaaavaggaagaaagaaagaaagaa 9461  
 QY 727 gtcccgcttctcaccgacagaagaagtgctgagtgagcc--cgtggagatgc 784  
 DB 9462 ggnhsrargvamtlthartrhrraasryshsngasahsryshsryshsraan 9521  
 QY 785 tgaagaagaagcgcgctggaatcagtgctgtgctgtagtcacacccctccaccact 844  
 DB 9522 vaasastrhsrhaasrstrasaagysraasysstrhsraaagstrtraaasraasgash 9581  
 QY 845 tcaagcaacagcaagcaacccagacagagagagagagagagagagagagagagagag 904  
 DB 9582 shargrgrysysgrsrycsysrvaayaasraasnmntaashvagtlysaavavas 9641  
 QY 905 acggagctcgtgctgagagcctgcccgaagaagaaacagtgagc-----gctatgaatg 959  
 DB 9642 rvagyaathrgnsrrsraasrysaavathrgasnsrgrthrygysgrthrrgsg 9701

QY 960 tcaaggcctgagctggacacacatgatacgtcgtgagtgaaacacagaaactactgct 1019  
 DB 9702 yysgrvarashsrygnthrrthrrvavaaasmgtmtsrashsraasryasy 9761  
 QY 1020 gaactatgtctcagcgtcagtgagtgcccgaccccttgagagagagagagagagag 1079  
 DB 9762 sasnthrcysrshsrygylthrvargtrgnaasvavaysthrngnaashvathrrhsy 9821  
 QY 1080 cctcaccttgacctgtgagcagagagtagccagagactcagtlccagtgctgagagaga 1139  
 DB 9822 yrasashasysvaaaarghvaasnaashashasrstrargysrvahsaathrgna 9881  
 QY 1140 agagacagcagcagtgctgcagaaagggcctgcttcagtgatgcatgacccaagcaga 1199  
 DB 9882 snrrgncysrshsrycsrstrasrsvavaasnrstraagygaasnaaasrysaargr 9941  
 QY 1200 ggcagagagcgact 1213  
 DB 9942 gngassysrsgyy 9955

# RESULT 9

ID Q04529 standard; DNA; 2072 BP.

AC Q04529;

DT 01-OCT-1990 (first entry)

DE Elastase gene.

KW Elastase; vaccine; ss.

OS Pseudomonas aeruginosa IPO 3455 strain.

FN Key Location/Qualifiers

FT CDS 363..1799

FT sig\_peptide 363..431

FT mat\_peptide 974..1796

FT /tag= c

FT /product=elastase

FT RBS 347..352

FT /\*tag= d

PN JP02104285-A.

PD 17-APR-1990.

PE 26-AUG-1988; 88JP-0213308.

PR 26-AUG-1988; 88JP-0213308.

PA (SUNR ) SUNTORY LTD.

DR WPI; 1990-161272/21.

DR P-PSDB; R04881.

PT Pseudomonas aeruginosa-derived elastase coding DNA -

PT in which chromosomal DNA of P. aeruginosa is cut into fragments,

CC inserted into E. coli vector for transformation, etc.  
 PS Claim 1; Fig 3; 13pp; Japanese.  
 CC The sequence encodes a protein with a signal peptide which is  
 CC cleaved to produce pro-elastase. this is further cleaved to give  
 CC the mature enzyme. The DNA can be used to produce the protein by  
 CC recombinant techniques for use in vaccines for P. aeruginosa infection.  
 XX  
 SQ Sequence 2072 BP; 424 A; 676 C; 623 G; 349 T; 0 other;



PD 16-FEB-1995



Query Match	2.2%	Score 42;	DB 16;	Length 6725;
Best Local Similarity	50.5%;	Pred. No. 0.65;		
Matches 102;	Conservative	0;	Mismatches 100;	Indels 0;
			Gaps	0
Yy 705 gaagagatccaggaagatccagcgtccctgtttctacccgcagacaaagtgtgctgga	764			
Dd 2541 gaagagagagaagaagctgtgcagactgcagcccgagagaagaacaaagatgtgtgga	2600			
Yy 765 agtggagcccggtgggaatgctgaaggaagggagacgggtggaatcaggtgttgcgtga	824			
Dd 2601 gaagccgagcagtggtgggaatccaaaggaggaagatgtgactgaaatccatccacggtga	2660			
Yy 825 tggcagaccctccaccacacttcagctcagcaagcagaccccgacccagggagggcaga	884			
Dd 2661 cggagagctctccaccgcgcacacagatcaacatgtgtgactccagcccaatggaatga	2720			
Yy 885 ggaagagacaacccaagcagacac	906			
Dd 2721 ggataagagccctacccaac	2742			
-RESULT 14				
ID X84332	X84332 standard; DNA: 5059 BP.			
AC X84332:				
DT 08-SEP-1999	(first entry)			
DE Stealth virus nucleic acid clone, SEQ ID NO: 24.				
DE Stealth virus; detection; diagnosis; infection; ss.				
OS Stealth virus.				
XX key	Location/Qualifiers			
XX misc-difference	3605			

FT	/tag=	a
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3610
FT	/tag=	b
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3615
FT	/tag=	c
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3630
FT	/tag=	d
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3631
FT	/tag=	e
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3638
FT	/tag=	f
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3641
FT	/tag=	g
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3642
FT	/tag=	h
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3657
FT	/tag=	i
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3659
FT	/tag=	j
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3681
FT	/tag=	k
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3698
FT	/tag=	l
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3703
FT	/tag=	m
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3708
FT	/tag=	n
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference	3713
FT	/tag=	o
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the

FT	misc_difference	3720	/*tag=	P	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3732	/*tag=	Q	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3733	/*tag=	R	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3740	/*tag=	S	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3747	/*tag=	T	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3750	/*tag=	U	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3751	/*tag=	V	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3766	/*tag=	W	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3778	/*tag=	X	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3792	/*tag=	Y	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3799	/*tag=	Z	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3800	/*tag=	aa	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3812	/*tag=	ab	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3813	/*tag=	ac	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3818	/*tag=	ad	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		

FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
FT		
misc_difference	3819	
FT	/tag=	ae
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3824	
FT	/tag=	af
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3827	
FT	/tag=	ag
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3841	
FT	/tag=	ah
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3853	
FT	/tag=	ai
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3857	
FT	/tag=	aj
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3860	
FT	/tag=	ak
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT		
FT		
misc_difference	3863	
FT	/tag=	al

Query Match	2.18;	Score 41.2;	DB 20;	Length 5059;
Best Local Similarity	17.0%;	Pred. No. 0.92;		
Matches 36;	Conservative 123;	Mismatches 149;	Indels 2;	Gaps 1
QY 144	cgagctctccagctcccaaggaacctccagccatcgcagctgtttctgcgcacaga	203		
DB 145	csmmmrnsccanradcatrgrvrsybdsrlryngncdcstrlttrsaargnyrssym	204		
QY 204	gaagcgagcatctattcttcggtgtgcgcagagccagagccagagcgaacctgagga	263		
DB 205	tmaanaetrsmctmacyccordnrgthlurdrldgnratamarycarmmrnsandc	264		
QY 264	cgaagcgagctcagagctccagagacagagggactc--tgagctgactaagtcac	321		
DB 265	anasbbyrdctcarcmdamagncldvranctcdsnasayvrnctscancctthmn	324		
QY 322	ccccagagagcgatctctgtgcccagggcgaagcgccctggtcccaagaatcgc	381		
DB 325	syslmsmdardmasasrlmmlatryclmthraycngclntctwcthammmdncly	384		
QY 382	atccagctccgcgtctacaaagctccggagagcacaacatccagctcaacccttggc	441		
DB 385	rhrvscgntcdctsnmmcmhncrclanrvrsayasnrlrrwhmmngcadnscthy	444		
QY 442	atccctgtgaacgaagaagcgccctgagag	471		
DB 445	hscmhchansmasadwngatcnthrrshs	474		

```
Search completed: March 26, 2001, 07:37:16
Job time: 3682 sec
```

XX	
AC	N93196;
XX	
DT	28-JUN-1990 (first entry)
XX	
DE	Monkey recombinant kallikrein gene.
XX	
KW	Monkey; kallikrein gene; vasodilator; male infertility;
XX	
OS	Monkey.
XX	
FH	Key
FT	misc_feature
FT	39..
FT	/*tag= a
FT	/note="start of monkey kallikrien sequence"
FT	78..848
FT	/*tag= b
FT	/note="mature protein"
XX	
PN	EP297913-A.
XX	
PD	04-JAN-1989.
XX	
PF	30-JUN-1988; 88EP-0306039.
XX	
PR	30-JUN-1987; 87US-0068594.
XX	
PA	(AMGE-) AMGEN INC.
XX	
PI	Lin FK, Lu HS;
XX	
DR	WPI; 1989-009139/02.
DR	P-PSDB; P93722.
XX	
PT	New recombinant kallikrein polypeptide(s) and encoding DNA -
PT	useful as vasodilators and for treating male infertility.
XX	
ES	Table IV; page 11; 43p; English.
XX	
CC	The monkey kallikrein gene is 93% homologous to that of the
CC	human genomic kallikrein gene. See also P92314, N93194,
CC	N93195, N93193, N93197, N93198, P93719, P93720, P93721, and
CC	P93722.
XX	
SO	Sequence 925 BP; 209 A; 281 C; 243 G; 192 T; 0 other;



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 05:41:24 ; Search time 1105.8 seconds  
(without alignments)  
12357.208 Million cell updates/sec

Title: US-09-653-961-1  
Perfect score: 1950  
Sequence: 1 atggggctccagctggt.....tgagcattagcccggaat 1950

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp14:\*  
 121: em\_estro15:\*  
 122: em\_estro16:\*  
 123: em\_estro17:\*  
 124: em\_estro18:\*  
 125: em\_estro19:\*  
 126: gb\_est58:\*  
 127: gb\_est59:\*  
 128: gb\_est60:\*  
 129: gb\_est61:\*  
 130: gb\_est62:\*  
 131: gb\_est63:\*  
 132: gb\_est64:\*  
 133: gb\_est65:\*  
 134: gb\_est66:\*  
 135: gb\_est67:\*  
 136: gb\_est68:\*  
 137: gb\_est69:\*  
 138: gb\_est70:\*  
 139: gb\_est71:\*  
 140: gb\_est72:\*  
 141: gb\_est73:\*  
 142: gb\_est74:\*  
 143: gb\_est75:\*  
 144: gb\_est76:\*  
 145: gb\_est77:\*  
 146: gb\_est78:\*  
 147: gb\_est79:\*  
 148: gb\_est80:\*  
 149: gb\_est81:\*  
 150: gb\_est82:\*  
 151: gb\_est83:\*  
 152: gb\_est84:\*  
 153: gb\_est85:\*  
 154: gb\_est86:\*  
 155: gb\_est87:\*  
 156: gb\_est88:\*  
 157: gb\_est89:\*  
 158: gb\_est90:\*  
 159: gb\_est91:\*  
 160: gb\_est92:\*  
 161: gb\_est93:\*  
 162: gb\_est94:\*  
 163: gb\_est95:\*  
 164: gb\_est96:\*  
 165: gb\_est97:\*  
 166: gb\_est98:\*  
 167: gb\_est99:\*  
 168: gb\_est100:\*  
 169: gb\_est101:\*  
 170: gb\_est102:\*  
 171: gb\_est103:\*  
 172: gb\_est104:\*  
 173: gb\_est105:\*  
 174: gb\_est106:\*  
 175: gb\_est107:\*  
 176: gb\_est108:\*  
 177: gb\_est109:\*  
 178: gb\_est110:\*  
 179: gb\_est111:\*  
 180: gb\_est112:\*  
 181: gb\_est113:\*  
 182: gb\_est114:\*  
 183: gb\_est115:\*  
 184: gb\_est116:\*  
 185: gb\_est117:\*  
 186: gb\_est118:\*  
 187: gb\_est119:\*  
 188: gb\_est120:\*  
 189: gb\_est121:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	651	33.4	690	135	BE786708 601475064
2	624.6	32.0	1034	110	BE619435 601473262
3	609.4	31.3	975	137	BE879940 601491164
4	603.8	31.0	637	105	BE257461 601111413
5	596.2	30.6	669	28	AL042548 DKEP4341
6	551.2	28.3	679	106	BE275425 601121423
7	532.6	27.3	553	136	BE871209 601448791
8	513.8	26.3	594	136	BE813148 RCI-BM000
9	466.4	23.9	715	135	BE786166 601474277
10	461.2	23.7	729	14	AA983120 ua35d07.r
11	442.6	22.7	564	105	BE234610 141772 MA
12	435.2	22.3	602	105	BE226518 ia21907.Y
13	422.6	21.7	642	93	AM611108 un77a02.Y
14	410	21.0	414	142	M78206 EST01799 Su
15	399	20.5	423	3	AA191148 zpb6e01.r
16	382	19.6	527	2	AA088962 mm49e11.r
17	371.8	19.1	382	5	AA325490 EST28484
18	368.2	18.9	563	8	AA522327 v145d03.r
19	337.6	17.3	528	21	AI528914 ud42h04.Y
20	319.2	16.4	403	5	AA323568 EST26450
21	317.6	16.3	468	19	AI325641 mm49e11.Y
22	311.8	16.0	403	108	BE484000 170309 BA
23	303.8	15.6	402	2	AA085691 zn63d06.r
24	302.8	15.5	442	87	AA231129 uc39h07.Y
25	300	15.4	433	108	BE482927 168887 BA
26	292.4	15.0	315	5	AA322791 EST25565
27	289.6	14.9	502	18	AI283948 q171g04.x
28	285.8	14.7	402	108	BE487749 176829 BA
29	279.2	14.3	388	136	BE863695 UTM-BHO-
30	272.8	14.0	353	23	AI664124 ue82b06.r
31	271.4	13.9	353	108	BE485906 173198 BA
32	268	13.7	532	136	BE851455 uw96d03.Y
33	267.8	13.7	313	90	AM392972 CMI-ST027
34	264.8	13.6	268	26	AI902799 RC-BT015-
35	264.2	13.5	269	26	AI902803 RC-BT015-
36	261.2	13.4	306	133	BE006161 RCO-BN012
37	255.6	13.1	272	6	AA349269 EST55868
38	236.2	12.1	262	97	AA997154 RC2-BN004
39	225	11.5	466	14	AA989790 uc79b03.Y
40	220.2	11.3	465	142	N39748 vx29h08.r1
41	216.8	11.1	340	105	BE226755 us71p08.Y
42	216.8	11.1	450	142	N31224 yx53d05.r1
43	213	10.9	512	105	BE257665 601113092
44	212	10.9	254	3	AA187703 zp72b12.r
45	208.4	10.7	237	136	BE838197 CMI-FN010

## ALIGNMENTS

RESULT 1  
 BE786708 LOCUS BE786708 690 bp mRNA  
 DEFINITION 601475064F1 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:3678246 5',  
 mRNA sequence.  
 ACCESSION BE786708  
 VERSION BE786708.1 GI:10207906  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 630)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: DCTD/DRP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM628 row: o column: 15  
High quality sequence stop: 673.  
Location/Qualifiers  
1..690  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3878246"  
/clone\_1ib="NIH\_MGC\_68"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."  
BASE COUNT 168 a 175 c 228 g 119 t  
ORIGIN

Query Match 33.4%; Score 651; DB 135; Length 690;  
Best Local Similarity 98.1%; Pred. No. 6.2e-153;  
Matches 680; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY 681 cgggctgcccgaatggaaccacataaagaagtcacgggaagtcacgcctgttttcta 740  
Db 1 cgggctgcccgaatggaaccacataaagaagtcacgggaagtcacgcctgttttcta 59  
QY 741 cccgaacaagaagtcgtgctgaagtcgaagccctggaatgctcgaaggaaggaagcgcg 800  
Db 60 cccgaacaagaagtcgtgctgaagtcgaagccctggaatgctcgaaggaaggaagcgcg 119  
QY 801 cgtggaatcaggtgttggctgcatggaacccctccaccaactcagcatcgaagca 860  
Db 120 cgtggaatcaggtgttggctgcatggaacccctccaccaactcagcatcgaagca 179  
QY 861 gaaccccgaccaggaaggaaggaaggaagcaacgaacgaagcggtccgtggtc 920  
Db 180 gaaccccgaccaggaaggaaggaaggaagcaacgaacgaagcggtccgtggtc 239  
QY 921 gaagcctgcccgaaggaacacagtcgtgcatgctcaatgctcagggccctggaactggaac 980  
Db 240 gaagcctgcccgaaggaacacagtcgtgcatgctcaatgctcagggccctggaactggaac 299  
QY 981 catgatalcgtctgcatggaacacaggaactactggtgaactatgctgacgtcgcg 1040  
Db 300 catgatalcgtctgcatggaacacaggaactactggtgaactatgctgacgtcgcg 359  
QY 1041 agtgaatcccgagccctggaagacaggaaggaagcagcgtcgaactgtgagc 1100  
Db 360 agtgaatcccgagccctggaagacaggaaggaagcagcgtcgaactgtgagc 419  
QY 1101 aagaagtaagcaggaagcctcgaatcccaatgctggaagaaaggaagcagaagtcgtga 1160  
Db 420 aagaagtaagcaggaagcctcgaatcccaatgctggaagaaaggaagcagaagtcgtga 479  
QY 1161 aaggggacctgtgctcagttgcatgacactgaacgggaaggaaggaagcagcatcgcgtg 1220  
Db 1161 aaggggacctgtgctcagttgcatgacactgaacgggaaggaaggaagcagcatcgcgtg 1220

Db 480 AAGGGGGCGCTGCTGCTTCAGTTCATGACCTGAACGGGAGGAGGAGGGCGCTATCGCTG 539  
QY 1221 cgtgagctgtgtccagcataccggactgaaccgacacagctgtgtaagctggccat 1280  
Db 540 CGTGCGCTGTGTGCTGCGCAGCATACCCGG--CGTAACCGACACACCTGTCTCAACGTGCCAT 597  
QY 1281 ttgtgcccccttgatgagatcgaaggaaggaaggttggtgtaagaagaatgtgt 1340  
Db 598 TTTGGCCCCCTTGATGAGGATTCAGAGAGGAGAGGAGGTGTGGGTGAAGAAGATATAGT 657  
QY 1341 gtgaatcgtctgtgtgaagcgtcagggcacc 1373  
Db 658 GTTGAATCTGTCTGTGAAGCGCTCGGGACCC 690

RESULT 2  
BE619435  
LOCUS BE619435 1034 bp mRNA EST 24-AUG-2000  
DEFINITION 601473262F1 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:3876255 5',  
mRNA sequence.  
ACCESSION BE619435  
VERSION BE619435.1 GI:9890373  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1034)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: DCTD/DRP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1CM623 row: 1 column: 16  
High quality sequence stop: 678.  
Location/Qualifiers  
1..1034  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3876255"  
/clone\_1ib="NIH\_MGC\_68"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.8 kb. Library constructed by Life  
Technologies."  
BASE COUNT 296 a 276 c 308 g 153 t 1 others  
ORIGIN

Query Match 32.0%; Score 624.6; DB 110; Length 1034;  
Best Local Similarity 97.7%; Pred. No. 2.9e-146;  
Matches 708; Conservative 0; Mismatches 9; Indels 8; Gaps 7;

QY 359 gccctggtccaggaagatccgcatccgcgtclacaaagtcggaaggaacca 418  
Db 1 GCCCTCGCTCCAGAGATACCGCATCCAGCTCCGCTTCAAAAGCTCCGGAGAGACCAA 60  
QY 419 acatccaggtcaaccccttgatccctgttgaaagtaaggaagccttgaggtgctga 478  
Db 61 ACATCCAGGTCAACCCCTTGATCCCTGTGTGAACGTAAAGAGCCTGAGAGGCTGCTTA 120  
QY 479 cctgtgaaggaaggaaggtaccatcctcaagtcacatcgtgtacaagaatggccg 538  
Db 479 cctgtgaaggaaggaaggtaccatcctcaagtcacatcgtgtacaagaatggccg 538

```
Db 121 CCTGTGATGAGGAGGAACGGGTACCCATTCTCAAGTCATCTGTATCAGAAATGCGCGC 180
QY 539 ccttgaagaaggagaagaacccgggtccacattcagtcgtcccaagactgtgagtcgagtg 598
Db 181 CTCTGAAGGAGGGAAGAACCGGGTCCACATTCAGTCGTCGCCAGACTGTGAGTCCAGTG 240
QY 599 gttctgaacactctgcaagactatctgaagcacagctgtgttaagaagaacaagatgccc 658
Db 241 GTTTGTACACTCTTGACAGATATTCTGAAGCACAGCTGGTTAAGAAGCAAAAGATGCC 300
QY 659 agtttactgtgaagctcaactacacggtctgcccagttggaaccacatgaagaagtcagg 718
Db 301 AGTTTACTGTGAGCTCAACTACCGGCTGCCAGTGGGAACCATGAAAGAGTCCAGGG 360
QY 719 aagtcacgtccctgtgttttctacccgacagaagaagtgtgctgtg -aagttgagccctg 777
Db 361 AACTCACCCGTCCTGTTTCTTACCCACAGAAAAGTGTGGCTGGCAAGTGGACCCGTTG 420
QY 778 ggaatctgaagaaggagccgctggaatcaggtgtttgtgtatgtgcaaccctcca 837
Db 421 GGAATGCTGAAGACCGGGACCCGCTGGAATCAGGTGTTGGCTGATGCAACCCCTCCA 480
QY 838 ccacactcagcaltcagcaagcaaaccccaagcaccaggaagcagaagaagacacc 897
Db 481 CCACACTTCAGCATCAGCAAGCAAGAACCCACAGCCAGGAGCAGAGAGAGACAA-C 539
QY 898 aagaagaacgggtgtcctgtgtgtgagccgtgcccgaagaagacagatgtggcgtatga 957
Db 540 AACGACACACGGGGTCTCTGGTGTGAGACCTG -CCGGAAGGACACAGTGGGCCCTATGAA 598
QY 958 tgtcagggcccttg -acttgaacacatgatcgtctgc -tgaagtgaacacacagaa -ct 1013
Db 599 TGTCAGGGCGCTGGGACTGTGGACACCATGATTCGCTGTGATGTAACACAGCAAGCT 658
QY 1014 actgtgtgaactatgtgtctgaagctccgagtgagtc -gacgccccttgagagcaagaag 1072
Db 659 ACTGTGTAACATATGTCTGACGTCGAGAGAGTCCCGGACCCCTGAGAGACCGGAG 718
QY 1073 gcagc 1077
Db 719 GCAGC 723

RESULT 3
BE879940 975 bp mRNA EST 27-SEP-2000
LOCUS BE879940 601491164F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3893607 5',
DEFINITION mRNA sequence.
ACCESSION BE879940
VERSION BE879940.1 GI:10328716
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 975)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTP/DRP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM668 row: o column: 16
High quality sequence stop: 645.
FEATURES
source Location/Qualifiers
1..975
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3893607"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pCMV-SpORT6; Site: 1; Note:
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 278 a 224 c 317 g 156 t
ORIGIN
Query Match 31.3% Score 609.4; DB 137; Length 975;
Best Local Similarity 95.8%; Pred. No. 1,9e-142;
Matches 723; Conservative 0; Mismatches 21; Indels 11; Gaps 9;

QY 549 ggaagaagaaccggttcacattcagtcgtcccaagactgtgagtcgagttgtgtacac 608
Db 1 GGAGAAAGAACGGGGTCCACATTTCAGTGTCTCCAGACTGTGAGTCGAGTGTGTACAC 60
QY 609 ctgcagaatattctgaagcagacagctgtgttaagaagaacaagatgccagtttactg 668
Db 61 CTTCGAGAGTATTCTGAAGGCACAGCTGTAAAGAAACAAATGCCCCAGTTTACTG 120
QY 669 tgagctcaactacccgctgcgccagtgaggaaacacatgaagaagtcaccaaggagtcacgt 728
Db 121 TGAGCTCAACTACCGGCTGTGCCAGTGGGAACCATATGAGAGATCCA -GGAAGTCACCGT 179
QY 729 cccgttttctaccccgacagaagaagtgtgtgtgaagtgtgagccgtgtggaatgtcgtgaa 788
Db 180 CCGTGTTTTTCACCCGACAGAAAAAGTGTGCTGGAAGTGAAGCCCTGGGAATGCTGAA 239
QY 789 ggaagaagacccgctgtggaatcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 848
Db 240 GGAAGGGGACCGCGTGGAATCAGGTGTTGGCTGTGATGGCAACCTCCACACACTTCAG 299
QY 849 catcagaacagaaaccccaagcaagcagaaggagca -gaggaagaagacaacaacagacaag 907
Db 300 CATCAGCAAGCAAGAACCCACAGCACCAGGAGGCACAGAGAAAGACAAACACACAAACG 359
QY 908 gggctctgtgtgtgagcctgtcccgaaagaaacaagtgaggcgtatgaatgtcaagggcc 967
Db 360 GGGTCTGTGCTGTGAGGCTGCCCGAGGAACACAGTGGCGCCTATGAAATGATCAGGGCC 419
QY 968 tggacttgacaaccatgtatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1027
Db 420 TGGACTTGTGACACCAATGATTCGTGTGTGAGTGAACCAAGAACTACTGTGAACTATG 479
QY 1028 tgtctgaagctccgaatgaagtcaccgagcccttgagaagacagaaggcagcagctcaccc 1087
Db 480 TGTCTGAGCTGCGAATGAGTGTCCGCGAG -CCCTGAGAGACAGAAAGGACAGCCTCAACC 538
QY 1088 tgaactgtgaagcagagaatgaagcagcctcagatltccagtgagctgagagaagaagacag 1147
Db 539 TGAACCTGTGAGGACAGAGAGTGTGACCGAGACCTCGAG -TCCACTGCTGTGAGAAAGACAG 597
QY 1148 gcccaagtctggaagaaggggcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1207
Db 598 GCCAGGTGTGAAGAGGGG -CTGTGCTTCATGTTGCAATGA -CTGAACGGAGGACAGGAG 655
QY 1208 gggagctatgcctgcgtgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1265
Db 656 GCGG -TATCGGTAGGGGTGGGTTTGTGCCAAGATTAACCGGGGCGGAAACAAAGACT 714
QY 1266 gtccaagtgccatttctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1300
Db 715 GGTACACCGTGC -ATATTGGCCCCCTTGGGTTG 747

RESULT 4
BE257461
```



LOCUS BE257461 637 bp mRNA EST 13-JUL-2000  
DEFINITION 60111413F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:3352034 5',  
mRNA sequence.  
ACCESSION BE257461  
VERSION BE257461.1 GI:9127936  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 637)  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
Plate: L10M150 row: j column: 03  
High quality sequence stop: 601.  
Location/Qualifiers  
1..637  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3352034"  
/clone\_1ib="NIH\_MGC\_16"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Library constructed by Ling Hong  
in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH-MGC library."  
BASE COUNT 125 a 195 c 199 g 118 t  
ORIGIN  
Query Match 31.0%; Score 603.8; DB 105; Length 637;  
Best Local Similarity 99.4%; Pred. No. 4,4e-141;  
Matches 627; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 361 cctcgtcccaagagatccagatccagctccgcgtctctacaaagctccgagagcccaac 420  
|||||  
DB 361 CCTCGCTCCCAAGAGATCCAGCATCCAGCTCCGCGTCTACAAAGCTCCGAGGCCAAMC 420  
QY 421 atccaggtcaacccttgatccctgtgaaacagtaagagcctgagaggtcgtacc 480  
|||||  
DB 421 ATCCAGGTCAACCCCTCGGCATCCCTGTGAACAGTAAGACACCTGAGAGGTCCGTACC 480  
QY 481 tcttgtagaggaagagagggagaccatctcctcaagatcgtgttacaagaatggcggct 540  
|||||  
DB 481 TGTGTAGGAGAGAACCGGATGCCCATTCCTCAAGTATGTGTACAGATGGCGGCTT 540  
QY 541 ctgaagagagagaaagaccgggt-ccacatcagtcgtctccagactgtgagtcagtg 599  
|||||  
DB 541 CTGAAGAGAGAGAAAGACCGGGTCCACATTCAGTGTCCAGACTGTGAGTCCAGTGG 600  
QY 600 ttgtacacctgacagatattctgaagca 630  
|||||  
DB 601 TTGTG-ACACCTTGACAGATATTCTGAAGCCA 630

## RESULT 5

AL042548 669 bp mRNA EST 29-FEB-2000  
LOCUS AL042548  
DEFINITION DKFZp43412321\_r1 434 (synonym: htes3) Homo sapiens cDNA clone  
DKFZp43412321 5', mRNA sequence.  
ACCESSION AL042548  
VERSION AL042548.1 GI:5422003  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 669)  
Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
TITLE EST (Blum, et al.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Blum H  
MIPS  
Am Klopfersplitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by LMU (Ludwig Maximilians University,  
Munich/Germany) within the cDNA sequencing consortium of the German  
Genome Project.  
No si sequence available.  
This clone (DKFZp43412321) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).

## FEATURES

source

1..669  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp43412321"  
/clone\_1ib="434 (synonym: htes3)"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

## BASE COUNT

159 a 197 c 182 g 130 t 1 others

## ORIGIN

Query Match 30.6%; Score 596.2; DB 28; Length 669;  
Best Local Similarity 99.0%; Pred. No. 3.6e-139;  
Matches 620; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1322 gggtagaagaagaatagtgtgtaacatctgtgtgaagcgtcaaggacaccccgagccca 1381  
|||||  
DB 1 GGGTGAAGAGAAATATGATGTGATCTGTGTGAAGCGTCAAGGACACCCCGGCCCA 60  
QY 1382 ccattcctgtaagcgtcaacgagcagtgaaacaagaaccaagattccacagcgatcc 1441

```
|||||
Db 61 CCAATCTCTGGAACGTCAAGCGACGCAAGTAAACAGCCAAAGTCCACACGCGATCC 120
QY 1442 tggagcaaccctgaatctgctcgtgaccccgagagctgcttgagagacagtgatitgaatgacag 1501
Db 121 TGAGACACCTGGAATGTCTCTGTGATCCCGGAGCTGTTGGACAGAGTGTGAATGACAGG 180
QY 1502 cctccaaagcagcttgycgaataaacacacagcaaccctctccttgagcgtgcaatllaacca 1561
Db 181 CCTCAACAGACCTGGGCAAAAACACAGCATCTCTTCTCGAGAGCTGTCAATTTAACCA 240
QY 1562 cctctaacacagactccacaacaacaacactgagctcagcaactccactgccaagtctcata 1621
Db 241 CCCTCAACACAGACTCCAAACACACACACTGGCCTCAGACCTTCCACTGCGACGCTCATATA 300
QY 1622 ccaagacccaacagcactccacacagagaagaagctgcccgaagcccgagagagccggag -cgtg 1680
Db 301 CCGAGAGCCACACACCTCCACAGAGAGAAAGCTGCCGGAGCCGGAGAGCCGGGGCCGTG 360
QY 1681 gtacatcgtgagctgtagtctgtgcatccctgctcgtgagcgtgctgagcgtgtccctat 1740
Db 361 GTCATGCTGGCTGTGATGTGTGTCATCTGTGCTGCTGCGGCTGTGGCGCTGTCTCTAT 420
QY 1741 ttcctctataagaagagcagagctgcccgtgcaagcgtcagggagcagagagatcacgctg 1800
Db 421 TTCCTCTATTAAGAAAGGCAAGCTGCCGTGACAGCGCTCAGGGAAGAGAGATCACGCTG 480
QY 1801 ccccgctctcgtgaagagcagactgtgagtgaagttgaatcagataagctcccaagaagag 1860
Db 481 CCCCCGCTCTGTAAGAGAGCAACTTGTAGTTGAAGTTAAGTCAGATTAAGCTCCCAAGAG 540
QY 1861 atggagctctcgtcagagcagcagcagctgacaaagaagagc -tccggagagcagagagagaa 1919
Db 541 ATGGAGCTCTCTGCAAGGACACACGCGGTGACAAAGAGGGCTTCCGGAGACACAGAGAGAA 600
QY 1920 atatcatcgatcgtgagcattagcccc 1945
Db 601 ATATCATCGATCTGAGGCAATTAGCCC 626

RESULT 6
LOCUS BE275425 679 bp mRNA EST 13-JUL-2000
DEFINITION 601121423F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2988868 5',
ACCESSION BE275425 mRNA sequence.
VERSION BE275425.1 GI:9150379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 679)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCID/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM78 row: n column: 05
High quality sequence stop: 588.
location/Qualifiers
1..679
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2988868"
/clone_lib="NIH_MGC_20"
```

```
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOT87; Site:1; XhoI: Site_2;
EcoRI: cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'-
adaptor: GGCAGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 159 a 196 c 197 g 127 t

ORIGIN

Query Match 28.3%; Score 551.2; DB 106; Length 679;  
Best Local Similarity 96.1%; Pred. No. 7.2e-128;  
Matches 640; Conservative 0; Mismatches 18; Indels 8; Gaps 7;

```
QY 143 gcgagctctccagtcaccagcaacctcaagcatalgtc-gactggttttc-tgtccacaa 200
Db 1 GCGGCTCTCCACAGTCCCAAGCAACCTCAGCAGATGCTGACTGTTTCTTGTCCACAA 60
QY 201 ggaagaagcagagcgtcattctccgtgtgctgcccgaagggccca-gagcgaacctggg 259
Db 61 GGAGAAAGCGAGGCTCATCTTCCGTGTGCGCCAGGGCCAGGGCCATGAGCGAACCTGGGG 120
QY 260 agtaagagcagcagctcagctccacagagaggggtactcgtgcccctgaactca 319
Db 121 AGTAGACAGAGCGGTGACGCTCCAGGACAGAGGGGCTACTCTGTGCTTACTTAATCA 180
QY 320 ccccccagaagcagcatalcttctgtgctcagggcaagcgccctcgtgtcccaagagtaac 379
Db 181 CCCCCAAGAGAGAGCGGCTATTCTTTGTGTCAGGGCAAGCGCCCTGGTCCAGAGATGACC 240
QY 380 gcatcagctcccggtctcaaaaagctccggagggccaaaatccag-gtcaacccctg 438
Db 241 GCATCAGCTCCGCGCTCTACAAAGCTCCGGAGGAGCCAAACATCATGTCAACCCCTTG 300
QY 439 -ggatccctgtgaacagtaagagccctgagagctgactacctgtgtaggaagagc 497
Db 301 TGCGATCCCTGTGAACAGTAAAGAGCCGTGAGAGGTCCCTACTCTGTGTAGGAGAGACGG 360
QY 498 gtaccatctcctcaagtcattcgtgtacaagaatgagccgctctgaagagagagaa 557
Db 361 GTACCCCATTCCTCAAGCATCTGTGTACAAAGATGGCGGCTCTGAAGAGAGAGAA 420
QY 558 ccgggtccaatcattcagtcgtgtcccaagctgtgagtgagtggtgtgtacacctgtcag 617
Db 421 CCGGCTCCACATTCAGTGTGCTCCAGACTGTGAGTTCAGTGTGTGACACTTTCAGAG 480
QY 618 tatctgaagcagacagctggtgttaagaagaacaagaatgcccagtttactgtgagctcaa 677
Db 481 TATTGTGAAGGACACAGCTGTGTCAAAAGAACCAAGATGCCAGTGTACTGTAGCTCAA 540
QY 678 ctacc-ggctgcccagtgaggacacatgaaagagctcaggaagatcaacctgctgtt 736
Db 541 CTACCTGATGCTCCAGTGTGAACCAATGAAGGAGTCAAGGAGAGTCAACCTACTGTGA 600
QY 737 tctaccgcagagaaaagtgtgctgtgagatgagagcccggtggaatgctgaagaaggg 796
Db 601 TCAACCCCAAGACAAAGTGTGCTTGAAGTg--AGCGTGGGCACTGTAAACAGACGGA 658
QY 797 accgag 802
Db 659 CGCGTG 664

RESULT 7
LOCUS BE871209 553 bp mRNA EST 27-SEP-2000
DEFINITION 60144791F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852671 5',
ACCESSION BE871209 mRNA sequence.
```

```

VERSION      BE871209.1  GI:10319985
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE    1 (bases 1 to 553)
AUTHORS      NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert_strausberg@nih.gov
              Tissue Procurement: ATCC
              CDNA Library Preparation: Life Technologies, Inc.
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM562 row: e column: 24
              High quality sequence stop: 551.
              Location/Qualifiers
                1..553
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3852671"
                /clone_lib="NIH_MGC_65"
                /tissue_type="adenocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI;
                Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.8 Kb. Library constructed by Life
                Technologies."
BASE COUNT   146 a 142 c 174 g 91 t
ORIGIN
Query Match      27.3%; Score 532.6; DB 136; Length 553;
Best Local Similarity 98.9%; Pred. No. 3.2e-123;
Matches 547; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 643 gaagaaagaatgcccagtttaattctgtgagctcaactaccggtccagtggaacac 702
    |||||||
DB 1 GAAGACAAAGATGCCAGTTTACTGTGAGCTCACTACCGGCTGCCAGTGGAAACAC 60

QY 703 atgaagagctcagggaatccacctccctgtttctaccacgaagaatgtgagctg 762
    |||||||
DB 61 ATGAAGAGATCCAGCAAGTCAACCGTCTGTCTTACCCGACGAAAGAAAGTGTGGCTG 120

QY 763 gaagtggagccgttggaaatgctgaagaaaggagccggtgaatcagtggttggct 822
    |||||||
DB 121 GAAGTGGAGCCCGTGGGATGCTGAAGGAAGGGGACCGGTGGAATCAGTGTGGCT 180

QY 823 gatgtgaacctccaccacattcagcatcagcaagcagaacccacagacaggaaggca 882
    |||||||
DB 181 GATGGAACCCCTCCACACTTTCAGCATCAAGCAAGAACCCACGACACCGAGAGCA 240

QY 883 gaggaagaacaacaagaagggtcctgtgtcgtggagcctgtcccggaagaacaac 942
    |||||||
DB 241 GAGGAAGAACACACACACACACGAGGCTGTGTGTGGACCTGTCCGGAAGGAAAC 300

QY 943 agtgggagctatgaatgtcaggagcttgacttgacacatatactgctgtgagtgaa 1002
    |||||||
DB 301 AGTGGGAGCTATGAATGTCAAGGGCTGGACCTTGACACCATCATATCTGCTGAGTAA 360

QY 1003 ccacaggaaactctgtgaactatgttctgaagtcgagtgagtcgccgagccctgag 1062
    |||||||
DB 361 CCACAGGAACCTACTGTGTAATGTGTGACGTCACAGTCAGTCCGCGAGCCCTGAG 420

QY 1063 agacaggaagcagagcctcacctcagcctgtgagcagagatgagcagaactcag 1122
    |||||||
DB 421 AGACAGGAAGCAGCAGCCTCAACTGACTGTGAGGACAGAGATGAGCAGACCTCGAG 480

```

```

QY 1123 ttccagtgcc--tgagagaagacagagccagtgctgtgaaagggtgtgttcacgt 1180
    |||||||
DB 481 TTCAGTGGCTCTGCAGAGAGAGACAGGCGGTCTGGAAAGGGGCGCTGTGCTTACGT 540

QY 1181 tgcatacctgaa 1193
    |||||||
DB 541 TCATATGACTGAA 553

RESULT      8
LOCUS       BE813148/C 594 bp mRNA EST 21-SEP-2000
DEFINITION RCI-BN0005-190700-017-h02 BN0005 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BE813148
VERSION    BE813148.1 GI:10245382
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE   1 (bases 1 to 594)
AUTHORS     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
             Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,R.F.,
             Goldmann,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
             Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
             ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
             Simpson,A.J.
             Shotgun sequencing of the human transcriptome with ORF expressed
             sequence tags
             Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
             Contact: Simpson A.J.G.
             Laboratory of Cancer Genetics
             Ludwig Institute for Cancer Research
             Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
             Brazil
             Tel: +55-11-2704922
             Fax: +55-11-2707001
             Email: asimpson@ludwig.org.br
             This sequence was derived from the FAPESP/LICR Human Cancer Genome
             Project. This entry can be seen in the following URL
             (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-RCI-BN0005-190
             700-017-h02&ts=2000-07-19&rl=1)
             Seq primer: puc 18 forward
             High quality sequence stop: 527.
             Location/Qualifiers
               1..594
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="BN0005"
               /dev_stage="Adult"
               /note="Organ: breast_normal; Vector: puc18; Site:1: SmaI;
               Site:2: SmaI; A mini-library was made by cloning products
               derived from ORESTES PCR (U.S. letters patent application
               No. 196,716 - Ludwig Institute for Cancer Research)
               profiles into the puc 18 vector. Reverse transcription of
               tissue mRNA and cDNA amplification were performed under
               low stringency conditions."
BASE COUNT   123 a 161 c 177 g 133 t
ORIGIN
Query Match      26.3%; Score 513.8; DB 136; Length 594;
Best Local Similarity 95.6%; Pred. No. 1.7e-118;
Matches 561; Conservative 0; Mismatches 22; Indels 4; Gaps 3;

QY 1284 tggccccccttgatgagcattcaagagagaggtgtgtgaagaataatagtgtt 1343
    |||||||
DB 594 TGGCCCCCTGGAGTGCATTCAAGAGAGATAGCTCTGGGTAAAGAAATATATTTGTT 535

QY 1344 gaatcttcttgaagcgtcagggacccccggccacacatctcctggaacgtcaag 1403
    |||||||

```

D	b	534	GATCTGTCTTGTGAAGCGTCAGGGGACCCCGGCCGCACCAACTCTCTGGAAGCTCTAACG	475
O	y	1404	cacggca-agtgaacaagaaccaaagatccaacgacgagtcttgagaccctgaatgtcctcg	1462
D	b	474	CACGGCATTAAGTATCAAGACCAAGATCCACAGGATCTCTGAGCACCCTGAAATGCTTCG	415
O	y	1463	ttagccccgagtcgtttggagaacgagttc--gaabcaacgagctcccaacgaacttgagaa	1520
D	b	414	TGACCCCCGAGCTGTTGGAGACAGTGTTTTGAATTGTCACAGGCTCTCCAAAGTCTGGGAT	355
O	y	1521	aaacacagatactctcttcctgagatctggtaaatbaaccacc-ctcacacagattcca	1579
D	b	354	TAAACACAGATCTCTCTCTCTGAGCTGGTCAATTAAACACACTCTCACACAGACTCG	295
O	y	1580	acacaacacctgagctcgaagcatcttcacatgcagtcceagtcctaataccagaaacagacct	1639
D	b	294	ACAACAACACTGGCGCGCAGCATCTTCCACTGCCCAGTCTCTATACAGAGCCAACACACTT	235
O	y	1640	ccacagagaagaagctctccggaacccgagaacccggaggctgtgtaactgtgtgatctg	1699
D	b	234	CCACAGAGAGATAGCTCTCCGAGCCGAGAGCCGGGGCGTGTCAATCGTGGCTGTGATTG	175
O	y	1700	tgtgacaccgagctctggtcggtgctgtggcgctgtgccctcatcttcclataaagaaggca	1759
D	b	174	TGTGCATCTGAGTCTGCTGGCGGTCTGGCGGCTCTCTATTATTCGTATAAGAAGGACA	115
O	y	1760	agctgcgcgtgcagagcgctcagaggaaagagagatacagctgtcccgcgctctgaagagcg	1819
D	b	114	AGCTGCGGTGCGAGCGCTCTCAGGGAGACAGAGATCTACCGTACCCCGCTCTCTAAGAGCG	55
O	y	1820	aactctgagtgaagttaagtcagataagctcccaagaagatatggc	1866
D	b	54	AAC TTGATTGTAAGTAACTCAGATTAAGTCTCCCAAGACGATGGCC	8
R	E	S		
L	O	C		
D	B	E	BE786166	715 bp mRNA EST 20-SEP-2000
D	F	I	601474277E1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3877338 5'	
A	C	S	mRNA sequence.	
V	E	R	BE786166	
K	E	N	BE786166.1 GI:10207364	
S	O		EST.	
O	R		human.	
O	R		Homo sapiens	
R	E	F	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
J	O		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
T	I		1 (bases 1 to 715)	
T	I		NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.	
J	O		National Institutes of Health, Mammalian Gene Collection (MGC)	
C	O		Unpublished (1999)	
C	O		Contact: Robert Strausberg, Ph.D.	
C	O		Tel: (301) 496-1550	
C	O		Email: Robert.Strausberg@nih.gov	
C	O		Tissue Procurement: DCTD/DRP/Gazdar	
C	O		CDNA Library Preparation: Life Technologies, Inc.	
C	O		CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
C	O		DNA Sequencing by: Incyte Genomics, Inc.	
C	O		Clone distribution: MGC clone distribution information can be	
C	O		found through the I.M.A.G.E. Consortium/LLNL at:	
C	O		http://image.llnl.gov	
C	O		plate: LLC626 row: 1 column: 19	
C	O		High quality sequence stop: 660.	

**FEATURES**  
**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:3877338
/clone_id="NH_MGC_68"
/clone_tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Lung; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT

```

BASE COUNT		Average insert size 1.8 kb.	Library constructed by Life Technologies.
ORIGIN	158 a 218 c 196 g 143 t		
Query Match	23.9%	Score 466.4	DB 135; Length 715;
Best Local Similarity	99.8%	Pred No. 1	5e-106;
Matches 467;	Conservative	0; Mismatches 1;	Indels 0; Gaps 0

OY	1448	cccgaaatgctcccgagaccgagcctcttgagacaagctttaa-gcagcgtccca	1507
Db	1	CCCTGAATGTCCTCGTAGACCCCGAGCTGTGGAGACAGGTTTAAATGCAAGGCGCTCA	60
OY	1508	acgaacctgggcaaaaaacacacagatccctctccctgagctggtcaattaacacacctca	1567
Db	61	ACGACCTCGGGCAAAAACACCAGCATCTCTCTCTGGAGCTGGCTCAATTTAAACACCCCTCA	120
OY	1568	caccagaactcaacaacaacacactgagctcagaactccacactccacttaaccagag	1627
Db	121	CACCAAGCTCCAAACACACACACTGGCTTCAGCACTTCCACTGCCAGTCTCATACCAAG	180
OY	1628	ccaaacagacctccacacagagaagaagctgcagagccgagagccgagcgctggtcatcg	1687
Db	181	CCAACACACACTCCACAGAGAGAAGCTGGCGGAGCCGGGAGCCGGGCGTGTCATCG	240
OY	1688	tggctgtgattgtgtgcataccctgctccctgacggtgcttgagcgctgctccctatttccct	1747
Db	241	TGGCTGTGATTGTGTGTGATCTGTGCTCTGGCGGTGTGGCGCTGTCTTAATTTCTCT	300
OY	1748	ataagaagagcaagctcgcgtgcagcgcgtccagagaagacagagatacagctgcgcccgt	1807
Db	301	ATAAGAAGGCGCACTCGCTCGTGCAAGGGCGCTCAGGAAACAGAGATCACGTAACCCCGT	360
OY	1808	ctcgtaagagcgaactgtgtgttgaaattgaatgcagataagctcccaagaagata-gggcc	1867
Db	361	CTCCTAAGAGGAACCTGTGATTGAATTAAATTAAGTCAATTAAGCTCCAGAAAGATGGGCC	420
OY	1868	tccctcagagcagcagcgtgtgaacaagaagcctccgagaaacacagagag	1915
Db	421	TCTCTCAGGCGACGCGGTGTGACAAAGGGCTCCGGGATACCAACGAGGAG	468

RESULT	10
LOCUS	AA983120
DEFINITION	729 bp mRNA EST 27-MAY-1998
ACCESSION	AA983120
VERSION	AA983120.1
KEYWORDS	GI:3160398
SOURCE	EST.
ORGANISM	house mouse.
REFERENCE	Mus musculus.
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 729)
TITLE	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelsting,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
JOURNAL	The Mashu-HMM Mouse EST Project
COMMENT	Unpublished (1996)
	Contact: Marra M/Mouse EST Project



|||||  
Db 200 GACTGTTTCTTCTCCACAGAGAACCCACATCTTCCGGGTGCGGAGGCCAG 259  
Oy 241 ggcagagcaaacctgaggagtaacagacagcgcctcagacagagggctact 300  
Db 260 ggcagagagagcctgagagagacagcgcctcagacagagggagcagct 319  
Oy 301 cggagcctgacacagtaacacacacagagagagcagctctctctgacagagagc 360  
Db 320 cggagcctgacacagtaacacacacagagagagcagctctctctgacagagc 379  
Oy 361 cctcag 420  
Db 380 ccccgctccag 439  
Oy 421 atccagagtaac 480  
Db 440 ATCCAGTCATTCCTTGGCATCTCTGTGAACAGTGAAGAGAGAGAGAGAGAG 499  
Oy 481 tctgtgag 540  
Db 500 TGTGTGGAG 559  
Oy 541 ctgaa 545  
Db 560 CTAAG 564

RESULT 12  
BE226518 602 bp mRNA EST 06-JUL-2000  
LOCUS BE226518 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA  
DEFINITION 5' similar to TR-095812 095812 CELL SURFACE GLYCOPROTEIN P1H12  
PRECUSOR .; mRNA sequence.

ACCESSION BE226518 GI:8931754  
VERSION BE226518  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 602)  
AUTHORS Wille, D., Meadows, A., Clifton, S., Hillier, L., Marra, M., Pape, D.,  
Mellon, D., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter,  
E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole,  
R., Tsagaris, H., Williams, T., Jackson, Y., and Bowers, Y.  
WashU-Harvard Pancreas EST Project  
Unpublished (2000)  
Other ESTs: 1a21907.x1  
Contact: Douglas Mellon  
WashU-Harvard Pancreas EST Project  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmellon@biohp.harvard.edu, meadows@fas.harvard.edu  
Library was constructed by Dr. Douglas Mellon DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Adimika Meadows  
(meadows@fas.harvard.edu)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 454.  
Location/Qualifiers

## FEATURES

Source  
1..602  
/organism="Mus musculus"  
/strain="ICR"  
/db\_xref="taxon:10090"  
/clone\_lib="Mouse E10 5 12 5 Pancreas cDNA Library"  
/sex="Both"  
/tissue\_type="Pancreatic Bud"  
/dev\_stage="Embryonic day 10.5 and 12.5, mixed"  
/lab\_host="DH10B"

/note="Organ: Pancreas; Vector: pSPOR1; Site\_1: Not 1;  
Site\_2: Sal I; Library constructed using Superscript  
Plasmid Library kit (Life Technologies). cDNA made by  
oligo-dT priming. Size-selected by column fractionation;  
average insert size 1.47 kb. Primary library, unamplified.  
cDNA Library Preparation: Goulin Chen."

BASE COUNT 157 a 166 c 147 g 132 t  
ORIGIN

Query Match 22.3%; Score 435.2; DB 105; Length 602;  
Best Local Similarity 82.8%; Pred. No. 9,7e-99;  
Matches 497; Conservative 0; Mismatches 103; Indels 0; Gaps 0;  
Oy 1339 gttgtgaatctgtctctgtgtgaagcgtcagagagaccccgccacacatctctgtgaagc 1398  
Db 3 GTGCTGANTCTGTCTTGTGTGAGGCTTCAAGACATCTCAGCCACATCTCTGTGAATGTC 62  
Oy 1399 aacgagcagcgaagtgacaagaacacacacacacacacacacacacacacacac 1458  
Db 63 AATGTTGCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 122  
Oy 1459 ctctgtgaccccgagagcgtgtgtgagagagagagagagagagagagagagagagag 1518  
Db 123 CTGTGAGCGCAGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 182  
Oy 1519 aaaaac 1578  
Db 183 TCMAAC 242  
Oy 1579 aacaac 1638  
Db 243 AGCCAAAC 302  
Oy 1639 tccacag 1698  
Db 303 TCCACAG 362  
Oy 1699 gttgtcaccctgtctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1758  
Db 363 GTGTGTACCTTGT 422  
Oy 1759 aagctgcgtcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1818  
Db 423 AAGCTGCATGT 482  
Oy 1819 gaactgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1878  
Db 483 GAATATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTAT 542  
Oy 1879 agcagcgttgacaagaagagcgtccgagagagagagagagagagagagagagagagag 1938  
Db 543 ATCAACTGTGACAAAGATGCTCCATGAGACAGGAGAGTAAATCATCATCATCATCATCAT 602

## RESULT 13

AW611108 642 bp mRNA EST 23-MAR-2000  
LOCUS AW611108  
DEFINITION un77402.y1 Sugano mouse kidney m1a Mus musculus cDNA clone  
IMAGE:2536874 5' similar to SW:W018.HUMAN P43121 CELLS SURFACE  
GLYCOPROTEIN MUC18 PRECURSOR .; mRNA sequence.

ACCESSION AW611108 GI:7315849  
VERSION AW611108  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 642)  
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wille, T.,  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wille, T.,  
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,  
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,  
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,

OY		744	gacagaaaaagtgttgccttggaagltggaaccgttgtgatcgtcaaggaaagggaccgcgt	800
Dd		1	GACAGAAAAGTGTTCCTGGAAGTGACCCCGTGGCATGCATCAAGAAGAGGGACC CGCT	600









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:37 ; Search time 19.45 seconds

(without alignments)  
1135.691 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGLPRLVCATLLAACCCCPR.....SSGDKRAPGDGEKIDLRH 646

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

268485

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*  
1: /SIDSL/gcgcdata/geneseq/AA1980.DAT:\*  
2: /SIDSL/gcgcdata/geneseq/AA1981.DAT:\*  
3: /SIDSL/gcgcdata/geneseq/AA1982.DAT:\*  
4: /SIDSL/gcgcdata/geneseq/AA1983.DAT:\*  
5: /SIDSL/gcgcdata/geneseq/AA1984.DAT:\*  
6: /SIDSL/gcgcdata/geneseq/AA1985.DAT:\*  
7: /SIDSL/gcgcdata/geneseq/AA1986.DAT:\*  
8: /SIDSL/gcgcdata/geneseq/AA1987.DAT:\*  
9: /SIDSL/gcgcdata/geneseq/AA1988.DAT:\*  
10: /SIDSL/gcgcdata/geneseq/AA1989.DAT:\*  
11: /SIDSL/gcgcdata/geneseq/AA1990.DAT:\*  
12: /SIDSL/gcgcdata/geneseq/AA1991.DAT:\*  
13: /SIDSL/gcgcdata/geneseq/AA1992.DAT:\*  
14: /SIDSL/gcgcdata/geneseq/AA1993.DAT:\*  
15: /SIDSL/gcgcdata/geneseq/AA1994.DAT:\*  
16: /SIDSL/gcgcdata/geneseq/AA1995.DAT:\*  
17: /SIDSL/gcgcdata/geneseq/AA1996.DAT:\*  
18: /SIDSL/gcgcdata/geneseq/AA1997.DAT:\*  
19: /SIDSL/gcgcdata/geneseq/AA1998.DAT:\*  
20: /SIDSL/gcgcdata/geneseq/AA1999.DAT:\*  
21: /SIDSL/gcgcdata/geneseq/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3341	99.3	646	20	Y42404
2	3327	98.9	646	20	Y42405
3	494.5	14.7	583	17	R97230
4	492.5	14.6	583	17	W06891
5	492.5	14.6	583	19	W47088
6	466	13.9	570	17	R97231
7	266	7.9	3117	21	Y53667
8	252	7.5	1380	20	Y08402
9	244.5	7.5	1381	20	Y13564
10	244.5	7.3	647	14	R39686
11	237.5	7.1	736	14	R39682
12	236	7.0	4412	21	Y53666

13	234	7.0	848	21	Y88565	Human NCAM 140KD 1
14	233	6.9	647	11	R08117	Vascular cell adhe
15	233	6.9	647	14	R38549	VCAM-6D. Syntheti
16	233	6.9	647	21	Y59498	Human VCAM-1 prote
17	233	6.9	1496	20	W81030	Melanoma associate
18	233	6.9	1496	20	Y70469	Human p53 target m
19	231.5	6.9	753	20	W83927	Human t85 protein.
20	231.5	6.9	828	12	R13905	Rabbit ATHERO-ELAM
21	230	6.8	735	14	R38550	VCAM/ICAM-1. Synt
22	230	6.8	739	12	R08118	Vascular cell adhe
23	230	6.8	739	13	R10316	1E7/267 staloglyco
24	230	6.8	739	14	R21081	VCAM variant with
25	230	6.8	739	15	R38548	VCAM-7D. Syntheti
26	230	6.8	739	15	R49733	Sequence of the 1E
27	230	6.8	739	19	W46734	Amino acid sequenc
28	228	6.8	702	10	P94014	Carinoembryonic c
29	227	6.7	739	12	R13906	Human VCAM-1. Hom
30	226	6.7	738	12	R13251	Human VCAM-1. Hom
31	226	6.7	1257	20	W14802	PECAM-1. Homo sap
32	225.5	6.7	1257	20	W74152	PECAM-1. Homo sap
33	225	6.7	698	16	R65168	Human L1 cell adhe
34	225	6.7	698	16	W22844	Carinoembryonic a
35	225	6.7	734	17	W00182	Human carinoembry
36	224.5	6.7	1070	18	W08747	Carinoembryonic a
37	224.5	6.7	1649	20	Y08404	Human colon carcin
38	224.5	6.7	1651	20	Y13566	Human Robo1 protei
39	224	6.7	702	10	P93999	Human Robo1 polyp
40	224	6.7	702	15	R54713	Amino acid sequenc
41	224	6.7	702	17	W06872	Carinoembryonic a
42	224	6.7	702	20	W83137	Carinoembryonic a
43	222.5	6.6	644	14	R39685	CEA protein. Homo
44	222	6.6	698	9	R81229	VCAM-6D/ICAM-2. S
45	222	6.6	702	9	P81222	Carinoembryonic a

## ALIGNMENTS

RESULT 1	
ID	Y42404 standard; Protein; 646 AA.
AC	Y42404;
XX	
DT	01-DEC-1999 (first entry)
XX	
DE	Human MUC18 amino acid sequence.
XX	
KW	prostate cancer; melanoma; cell adhesion; glycoprotein;
KM	metastasis; treatment; detection; diagnostic test.
XX	
OS	Homo sapiens.
XX	
PN	W09945392-A1.
XX	
PD	10-SEP-1999.
XX	
PE	02-MAR-1999; 99WC-US04850.
XX	
PR	03-MAR-1998; 98US-0076664.
XX	
PA	(UYEM-) UNIV EMORY.
XX	
PI	Wu G.
XX	
DR	NPI; 1999-540899/45.
XX	
DR	N-PSDB; Z20930.
XX	
PT	Drosophila sp. Rob
XX	
PT	Drosophila Robo 2
XX	
PS	VCAM-6D/ICAM-1.
XX	
PS	VCAM/ICAM-2. Synt
XX	
XX	Sequence gl/101742
XX	
XX	Claim 17; Page 7; 80pp; English.

CC This is the amino acid sequence of the Human MUC18 cell adhesion  
CC glycoprotein, which is expressed on the surface of melanoma cells.  
CC and can be used as a marker for prostate cancer.  
CC This MUC18 (hMUC18) cDNA sequence is different from the hMUC18 cDNA  
CC given in Genebank accession number N28882, 220931. The deduced amino acid  
CC residues from this cDNA are very different from that published by  
CC Johnson's group in 1989.  
CC The presence of this glycoprotein has been correlated with the ability of  
CC melanomas to metastasize. MUC18 is also associated with normal vascular  
CC tissue, and on the smooth muscle of venules, and it expresses  
CC sporadically on capillary epithelium.  
CC The method can be used as a diagnostic test for prostate cancer which  
CC has a relatively high potential for metastasis or which has metastasized.  
CC The physician can then choose the appropriate surgical, chemotherapeutic  
CC or radiation treatment regimens. In addition, antibody specific to MUC18  
CC can be used to prevent metastasis of PCCs.

XX  
XX  
SO Sequence 646 AA:

Query Match 99.3%; Score 3341; DB 20; Length 646;  
Best Local Similarity 99.5%; Pred. No. 2,76-227;  
Matches 643; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVEVGSSTALLKCGLSQSGNLSHV 60  
DB 1 mglplvcafliaacccprvagvgeaeqpaeiveevgsstallkcglsqsgnlsHV 60  
QY 61 DMFSVHKERTLIFRVROGOGSEPEGEORLSLDRGATLALQVPODERIFLCOGKR 120  
DB 61 dmfsvhkertlifrvgogsepegeorlsldrgataltqvpoderiflcgkr 120  
QY 121 PRSQEYRIQLRYKKAPEEPNIQVNLGIPVNSKEPEEVAATCGRNQYIPQVYWKNGRP 180  
DB 121 prsqeyriqlrykkapeepniqvnlgipvnskepeeavatcgrnqyipqviwkngrp 180  
QY 181 LKEEKNRHHIOSQVSSGGLYTLQSLIKAOVKEEDKDAOFYCELNRYLPSGNHMKESRE 240  
DB 181 lkeeknrhiosqvsessgyltqlslkqavlkedkdaofycelnrylpsgnhmkesre 240  
QY 241 VVVPVFPTEKWLVEPEVGMLEKEDRVEIRCLADGNPPHFSISKONPSTREAEEETT 300  
DB 241 vvvpvfptekwlvepevgmlkegdreveircladgnpphfsiskpnstreaeeett 300  
QY 301 DNGVNLVLEPAKREHSGRYECQGLDIDTMSILSEPOELLVNVSDVRVSPAPEROEGSS 360  
DB 301 dngvnlvleparkehsgryecqglididtmisilsepellvnyvsdvrvsapaerogess 360  
QY 361 LFTTCEASSQDLFEQWLREETGOVLERGPVLQMLDKREAGGYRCVASVPSIPGLNRT 420  
DB 361 lfttceassqdlfeqwlreetsgvlergpvlqlhdkreaggyrcvasvpsipglnrt 420  
QY 421 QLVNVAIFGPPWMAFERKRWAKENWVNLSCBASHPPTTSMWNGTASQDDPOKRV 480  
DB 421 qlvnvaifgppwmaferkrywvknwvnlscbasghprrttsmwngtasqddpqkrv 480  
QY 481 LSTLVNLYPELLETGVECTASNDLGKNTSILELVNLTTLTTPDSNTTGLSTASPH 540  
DB 481 lsltnvlypelletgvectasndlgkntsilelvnlttltpdsnttglstasph 540  
QY 541 TRANSTTERKRLPEPSRCGVIVAVTICILVAVLIGAVLYFLYKKGKLPGRSGKOETLR 600  
DB 541 transtterkrlpepsrgvivaavtictilvavligavilyflykkglpgrsgkoetlr 600  
QY 601 PPSRKSLVVEVSKDLPEMGILLOSGSGKRAPGOGKEYYTLDRH 646  
DB 601 ppsrkslvvevskdlpeemgilgssgdkrapgogkexytlrhn 646

RESULT 2  
Y42405  
ID Y42405 standard; Protein: 646 AA.  
XX

AC Y42405:  
XX  
DT 01-DEC-1999 (first entry)  
XX  
XX Human MUC18 amino acid sequence. (Johnson 1989).  
DE  
XX  
XX prostate cancer; melanoma; cell adhesion; glycoprotein;  
KW metastasis; treatment; detection; diagnostic test.  
OS  
XX Homo sapiens.  
XX  
XX WO945392-A1.  
XX  
XX 10-SEP-1999.  
PD  
XX  
XX 02-MAR-1999; 99WO-US04850.  
PF  
XX  
XX 03-MAR-1998; 98US-0076664.  
PR  
XX  
XX (UYEM-) UNIV EMORY.  
PA  
XX  
XX Wu G:  
PI  
XX  
XX WPI: 1999-540899/45.  
DR  
XX N-PSDB: 220931.  
PT  
XX Detection of metastatic prostate cancer, by detection of MUC18  
PS expression in prostate cancer cells -  
PS  
XX Disclosure; Page 8; 80pp; English.  
XX  
XX This is the amino acid sequence of the Human MUC18 cell adhesion  
CC glycoprotein, which is expressed on the surface of melanoma cells,  
CC and can be used as a marker for prostate cancer.  
CC This sequence was originally published in 1989 by Johnson et al  
CC (Genebank N28882).  
CC This sequence differs from the one proposed by the inventors, this amino  
CC acid sequence is 603 amino acid residues while the present sequence  
CC 220930 has 646 amino acid residues.  
CC The presence of this glycoprotein has been correlated with the ability of  
CC melanomas to metastasize. MUC18 is also associated with normal vascular  
CC tissue, and on the smooth muscle of venules, and it expresses  
CC sporadically on capillary epithelium.  
CC The method can be used as a diagnostic test for prostate cancer which  
CC has a relatively high potential for metastasis or which has metastasized.  
CC The physician can then choose the appropriate surgical, chemotherapeutic  
CC or radiation treatment regimens. In addition, antibody specific to MUC18  
CC can be used to prevent metastasis of PCCs.

XX  
XX  
SO Sequence 646 AA:

Query Match 98.9%; Score 3327; DB 20; Length 646;  
Best Local Similarity 98.9%; Pred. No. 2,66-226;  
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPLVCAFLIAACCCPRVAGVGEAEQAPAEVEVGSSTALLKCGLSQSGNLSHV 60  
DB 1 mglplvcafliaacccprvagvgeaeqpaeiveevgsstallkcglsqsgnlsHV 60  
QY 61 DMFSVHKERTLIFRVROGOGSEPEGEORLSLDRGATLALQVPODERIFLCOGKR 120  
DB 61 dmfsvhkertlifrvgogsepegeorlsldrgataltqvpoderiflcgkr 120  
QY 121 PRSQEYRIQLRYKKAPEEPNIQVNLGIPVNSKEPEEVAATCGRNQYIPQVYWKNGRP 180  
DB 121 prsqeyriqlrykkapeepniqvnlgipvnskepeeavatcgrnqyipqviwkngrp 180  
QY 181 LKEEKNRHHIOSQVSSGGLYTLQSLIKAOVKEEDKDAOFYCELNRYLPSGNHMKESRE 240  
DB 181 lkeeknrhiosqvsessgyltqlslkqavlkedkdaofycelnrylpsgnhmkesre 240  
QY 241 VVVPVFPTEKWLVEPEVGMLEKEDRVEIRCLADGNPPHFSISKONPSTREAEEETT 300

```

Db      241 vtvrvfyftekvwleevvgmlkegdvrelcladgnpphfsiskqnpstreaeectn 300
QY      301 DNGVLVEPARKHSGRYECQGLDLOMTISLSEPOELVNVYSDVRSPAPEROEGSS 360
Db      301 dngvlyleparkehsgryecqawltmtisllsepeellvnyvsdvrsapaerqegss 360
QY      361 LTLTCEAESSODLEFQWLREETQVLERGPVLOLHDLKREAGGVCVAVSPISPLNFT 420
Db      361 lltlceaessdlefqwlreectqvlergpvlqhlkreaaggvcvavspisplnft 420
QY      421 QLVNVALFQPPMAFKERKWKVAKENNVNLSCBASGHPPTISWNVNCTASBDQDPQRY 480
Db      421 qlvkvalffqppmafkerkwvkenmvnlscasghppltsmwngfasedqdpqrv 480
QY      481 LSTNLVLPPELLETCVECTASNDLCKNTSILELVNLTTLTPSDNTTGTSTASPH 540
Db      481 lstnlvlppelettecvectasndlgkntslllelvnlttltpdsnttltgstasph 540
QY      541 TRANSTSTERKLPESRQGVIVAVICILVAVLGAVALFYFKGKLPGRSGKOETL 600
Db      541 tranststerklpepsrgvivaivcillvavlgavalfyfkkgklpgrsgkoetl 600
QY      601 PPSRKELVVEKSDKLPPEMGLGSSGDKRAPGDGGERYIDLRRH 646
Db      601 psirkelvveksdklpeemglgssgdkrapgdggekryidlrh 646

RESULT 3
R97230
ID      R97230 standard; Protein: 583 AA.
XX
AC      R97230;
XX
DT      08-OCT-1996 (first entry)
XX
DE      Stem cell marker HCAPro.1.
XX
KW      Stem cell marker; haematopoietic cell antigen; HCA; HCASeq.1;
KM      HCAPro.1.
XX
OS      Homo sapiens.
XX
PN      EP16146-A2.
XX
PD      12-JUN-1996.
XX
PF      02-FEB-1995; 95EP-0300661.
XX
PR      06-DEC-1994; 94US-0352323.
XX
PA      (SYST-) SYSTEMIX INC.
XX
PI      Gearing D, Uchida N, Yang Z;
XX
DR      WPI: 1996-269990/28.
XX
DR      N-PSDB: T28819.
XX
PT      DNA encoding human haematopoietic cell antigen proteins - useful as
PT      stem cell marker proteins in functional studies and for antibody
PT      prodn.
XX
PS      Claim 8; Fig 4A; 17pp; English.
XX
CC      HCAPro.1 (R97230) and HCAPro.2 (R97231) are novel human
CC      haematopoietic cell antigens associated with stem cells. They
CC      are the respective products of nucleic acid sequences HCASeq.1
CC      (T28819) and HCASeq.2 (T28820) obtd. from human CD34+ bone marrow.
CC      Expression systems are provided for prodn. of recombinant HCA
CC      proteins. The proteins are useful as stem cell marker proteins in
CC      functional studies, and can also be used to produce antibodies
CC      that allow the purification of stem cells from haematopoietic and
CC      other sources.

```

```

XX      XX      Sequence      583 AA:
SQ      SQ
Query Match      14.7%; Score 494.5; DB 17; Length 583;
Best Local Similarity      24.8%; Pred. No. 6.5e-27;
Matches 145; Conservative 121; Mismatches 236; Indels 63; Gaps 23;

QY      41 GSTALLKCGLSQSGNLSHVDFSVYHKEKRTLLFRVRGOGQS----EPGEYEQRLSD 96
Db      36 gdtllpcrltdvpq-nlmlfgkwnyekpdpqspvllaftssktsqyddvpeyadrlhse 94
QY      97 RGATLALQVTPQDERITLCO-GKPPSQEYRIOLRYKAPAEENIOVNLGIPVNSKEP 155
Db      95 -nyltsisnarnsdckrtvcmlvtedhvfearlvkfkqpskpeivskalfi--eteql 151
QY      156 EEVATCVGRNGVPIPOVIVYKNGRPLKEEKRVHIIQSSQYVES-SGLYTLQSTLKALQV 214
Db      152 kklgdci sedsyppdgnltwyrngkvlnpllegavvllfkemdpvtqlylmtstleytk 211
QY      215 EDKDAQFYCELNRYLPSGNHMKESREYTVPYFTEKRWLEV-EPVGMLEKGRVETRCL 273
Db      212 adlqmpftcsvtylygpsqgkclhsegavfdlylpteqvltqvlppknaikgdnlitkl 271
QY      274 ADGNPPRH--FSISKQNPSTRAEETINDGVVLEPARKHSGRYECQGLDLOMTIS 330
Db      272 gngppppeeelflylpgpberlrsntytltd-----vyrnatgdykcslldkksmia 323
QY      331 LLESEPOELVNVYSDVRSPAPAR-EROEGSSLTPTCEAESSODLEFQWLREETQVLERG 389
Db      324 sta-----ltvhyldslpsgvevtrigdallpsclsisrnatlvvmkdnlr--lrss 376
QY      390 PVLOLHDLKREAGGVCVAVSPISPIGINTQLVNVAIFQPPMAFKERKWKVAKENNVIN 449
Db      377 p--sfsslhqdagnyccetalgveglkkrseitllivegkp--qlkmtkktidpslstk 432
QY      450 LSCBASGHPPTISWNVNCTAS---EODDQPO---RYLSTNLVLPPELLETCVECTASN 503
Db      433 lchvegfpkpaqlwtlqgsavlnqleespylngrtys--kllispeenvt-ltctaen 489
QY      504 DUGKNTSILELVNLTTLTPDSNTTGTSTASPHSTRANSTSTERKLPESRQGVIV 563
Db      490 qlertvsnlvsalsi-----pe-----hdeadeidsentr-ekvndgklliv 530
QY      564 AVIYCIIVLAVLGAVALFYLY-KKGKLPGRSGKOETLPPSRKSE 607
Db      531 givvglllaalvagvywlymkkskaskhvnkdlgmneenkkle 575

RESULT 4
W06891
ID      W06891 standard; Protein: 583 AA.
XX
AC      W06891;
XX
DT      18-MAR-1997 (first entry)
XX
DE      Human activated leukocyte-cell adhesion molecule ALCAM.
XX
KW      Activated leukocyte-cell adhesion molecule; ALCAM; CD6 ligand;
KW      thymic epithelial cell; inflammation; allograft rejection;
KW      neurodegenerative disease.
XX
OS      Homo sapiens.
XX
FH      Key      Location/Qualifiers
FH      Peptide      1..27
FT      /label= Sig-peptide
FT      Protein      28..583
FT      /label= Mat-protein
FT      Domain      28..527
FT      /label= Extracellular domain
FT      /note= "the extracellular domain can be subdivid"

```

FT	Domain	528..551
FT	/label= "Transmembrane_domain"	
FT	Domain	552..583
FT	/label= Extracellular_domain	
FT	Modified-site	91
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	95
FT	Modified-site	95
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	167
FT	Modified-site	167
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	258
FT	Misc-difference	258
FT	/label= Asn, Ser	
FT	/note= "variation owing to polymorphism in HLEO and T-cell derived clones"	
FT	Modified-site	265
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	301
FT	Misc-difference	301
FT	/label= Met, Thr	
FT	/note= "variation owing to polymorphism in HLEO and T-cell derived clones"	
FT	Modified-site	306
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	337
FT	Modified-site	337
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	361
FT	Modified-site	361
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	457
FT	Modified-site	457
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	480
FT	Modified-site	480
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	499
FT	Modified-site	499
FT	/label= Glycosylation	
FT	/note= "putative N-glycosylation site"	
PN	W09634880-A1.	
PD	07-NOV-1996.	
XX		
PF	29-APR-1996;	96WO-US06010.
XX		
PR	01-MAY-1995;	95US-0432016.
XX		
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
PA	(UYDU-) UNIV DUKE.	
PI	Aruffo A, Bowen MA, Haynes BF, Marguardt H, Patel D;	
DR	WPI: 1996-506097/50.	
XX	N-PsDB: T46075.	
PT	CD6 ligands present on surface of thymic epithelial cells - used to develop prods. for treating e.g. inflammation, organ allograft rejection or neuro-degenerative diseases	
XX		
PS	Example 4; Fig 29; 112pp; English.	
CC	Human activated leukocyte cell adhesion molecule (W06891) or ALCAM	
CC	is a CD6 ligand present on the surface of thymic epithelial cells,	
CC	monocytes, activated T-cells, and other cells. Its amino acid	
CC	sequence was deduced from cDNA clones (T46075) obtd. from HL60	
CC	and T-cell lines. ALCAM polypeptides can be produced, e.g. as a	
CC	fusion proteins, in transformed host cells. They are useful for	
CC	inhibiting CD6/CD6 interactions between cells and to screen for	
CC	agents that modulate this interaction. Inhibitors can be used to	

CC	trealt eg. inflammation, multiple sclerosis, inflammatory uveitis,
CC	rheumatoid arthritis, T-cell mediated vasculitis syndromes, organ
CC	allograft rejection and neurodegenerative diseases.
XX	
XX	Sequence 583 AA:
Query Match	14.6%: Score 492.5; DB: 17; Length 583;
Best Local Similarity	24.8%: Pred. No. 9e-27;
Matches 145; Conservative 121; Mismatches 25e; Indels 63; Gaps 23	
QY	41 GSTALTKCGLSOSQGNLSHVDWFVSHKEKRTLLFVRVGGOGOS---EPGEYDRLSLQD 96
DB	36 gdtlllprldvprq-nlmgfkwwkxekpdpfpfiatrstsksvqddvpeyktlnlsl 94
QY	97 RGATLALVQVTRQDEKIFLPCQ-GKRRSRSEYKQLRVYKAPKEPNQVAPLGPVNSKEP 155
DB	95 -nytlisnarisdetrftvcmlqvtednvaefaprlvkvfqpsspelvskalfi--eteql 151
QY	156 EEVATGVGNNGRPDPVIVYKNGRPKEEKNRNYLIQSOTVES-SGLVTLQSLTKAQLVK 214
DB	152 krlgdcisdedsydpqnlitvyrngkvlhrllegavvllfkkmprvtdqlymtstleyktek 211
QY	215 EDKDAQFYCELNRLPSGNHMKESREYTVPVRYPTKYNLEV-EPYGMILKEGDRVIRCL 273
DB	212 adlqmpftcscvtygspgqklthseqavfdlyprteqvllqvlpkxxalkqgnltlckcl 271
QY	274 ADGNPPPH---FESIKONSTREAREEETINDNGVLVLEPARKESHSGRYCCOGLDPTMS 330
DB	272 gmpnppeellfllpqppegirsnetylxd-----vrrnatgdycslldkksmla 323
QY	331 LISEPEQLLVNVYSDYRVSPAP-EROEGSSLTLTCEAESSQDLRFQWLREETGVLENG 389
DB	324 sta-----ltvhy-l-dislpsgevtvrlgdaipvscslsraatvymkdnlr--lrss 376
QY	390 PVLQLMDLKRKADGAGRCVAVSPISGLNLTQVAVAFICPPMAKEREKRVVWVKENMVIN 449
DB	377 P--sfslslyhgdqgnvycetaqlgeveglkkesrltlivegkp--qlkmtkktlpsglstk 432
QY	450 LSCSEASGHPRTISWVNVGTAS--EQDDDPQ---RVLSLTNLVLPPELLETGVECTASN 503
DB	433 lltchvegfprkpalqwtltlgsavlnqleespylngryys--klilspreenvt-ltctaen 489
QY	504 DLGKNLSLFLLELVNLTLLPDSNTTTLGSLTVASPHTRANSTSTEEKLUPESPESRGVIV 563
DB	490 qlsetvnslnvsalsl---pe-----hdeadeidsendr-ekvndgqkliv 530
QY	564 AVITICILVAVLGAIVLYFL-KKKGKLCPRKRSKQETILPPSRKSE 607
DB	531 glivglllaalvagvvywylmkkkskaskhvnkdlgmeenkkle 575
RESULT 5	
W47088	
ID	W47088 standard; Protein: 583 AA.
XX	W47088;
XX	20-JUL-1998 (first entry)
XX	Activated leukocyte-cell adhesion molecule (ALCAM).
XX	ALCAM: activated leukocyte-cell adhesion molecule; CD6 ligand;
XX	antibody; human.
XX	Homo sapiens.
XX	
XX	Key Location/Qualifiers
XX	FT Peptide 1..27
XX	FT Protein /label= Sig_peptide
XX	FT Protein 28..583
XX	FT Protein /label= Mat_protein
XX	FT Domain 28..527

FT		/note= "extracellular domain"
FT	Domain	528..351
FT		/note= "transmembrane domain"
FT	Domain	552..583
FT		/note= "cytoplasmic domain"
FT	Modified-site	91
FT		/note= "N-glycosylated"
FT	Modified-site	95
FT		/note= "N-glycosylated"
FT	Modified-site	167
FT		/note= "N-glycosylation"
FT	Misc-difference	258
FT		/label= Asn, Ser
FT		/note= "encoded by ARK (polymorphism in nucleotide sequence"
FT	Modified-site	265
FT		/note= "N-glycosylated"
FT	Misc-difference	301
FT		/label= Met, Thr
FT		/note= "encoded by AVG (polymorphism in nucleotide sequence)"
FT	Modified-site	306
FT		/note= "N-glycosylated"
FT	Modified-site	337
FT		/note= "N-glycosylated"
FT	Modified-site	361
FT		/note= "N-glycosylated"
FT	Misc-difference	447
FT		/note= "encoded by CAR"
FT	Modified-site	457
FT		/note= "N-glycosylated"
FT	Modified-site	480
FT		/note= "N-glycosylated"
FT	Modified-site	499
FT		/note= "N-glycosylated"
PN	W09803551-A1.	
PD	29-JAN-1998.	
XX		
PF	17-JUL-1997;	97WO-US12511.
XX		
PR	18-JUL-1996;	96US-0684594.
PA	(BRIM ) BRISTOL-MYERS SQUIBB CO.	
PA	(UYDU-) UNIV DUKE.	
PI	Anuffo A, Bowen MA, Haynes BF, Marguardt H, Patel D; Sladak AW;	
PI		
DR	WPI: 1998-120703/11.	
XX	N-PSDB: V13954.	
PRT	Activated leukocyte-cell adhesion molecule (ALCAM), a new CD6 ligand	
PRT	- useful for screening compounds which inhibit interaction of ALCAM	
PRT	and CD6 ligand antibodies	
XX		
PS	Claim 3; Fig 29; 130pp; English.	
CC	This polypeptide is a human CD6 ligand (see W47088), designated	
CC	activated leukocyte-cell adhesion molecule (ALCAM), that is present	
CC	on the surface of thymic epithelial cells, monocytes, activated	
CC	T cells and a variety of other cells. Its amino acid sequence was	
CC	deduced from cDNA clones (see V13954) isolated from human T cell	
CC	and HL60 cDNA libraries. It is the human homologue of chicken	
CC	BEB, and shows homology to neurotin, RAGE and MUC18. Anti-CD6	
CC	ligand antibodies, or their binding fragments, are useful for	
CC	inhibiting binding of CD6 present on the surface of a first cell	
CC	to that present on a second cell. CD6 ligands and anti-CD6 ligand	
CC	antibodies are also useful in screening test compounds for the	
CC	ability to inhibit binding of CD6 ligand to an anti-CD6 ligand	
CC	antibody (all claimed).	
XX		

Seq	Sequence	583 AA;
Qy	Query Match	14.6%; Score 492.5; DB 19; Length 583;
Db	Best local similarity	24.8%; Pred No.96-27;
Matches	145; Conservative	121; Mismatches 256; Indels 63; Gaps 23.
Qy	41 GSTALKCGISOSQGNLSHVDWFSVHKERKTLIFRVROGOGS---EPGEYERLSTOD	96
Db	36 gdtlllpcrdldypq-nlmfgkwyekpdxspflafstrstksvqdvdyekdrlnlse	94
Qy	97 RGATLALVQVTPDDEKRLPICO--GKRPSQGEYRIQLVYKAPREPNIQVNPGLIPVNSKSP	155
Db	95 -nytlisnarisderrfrcmlytedanvteaprlvkvfpqskpelvskalfi--eteql	151
Qy	156 EEVATCVGRNGRPDPVPIVYKNGRPLKEEKNVHIIOSQTVES--SGLYLOSILKAQIVK	214
Db	152 krlgdcisdedsydpdgnlityrngkvlnrliegavvllfkfemprvqglymtstlleykttk	211
Qy	215 EDKDAFYCELVNRLPSSGNHKESEKRVIVPVVYPIPEKVALEY-EPVGMILKEDRAVEIRCL	273
Db	212 adlqmpfctsvlygspsgqkllhsegavfdlylpteqvlylplpkxalkgednaltlckl	271
Qy	274 ADGNPPPH---FSISKONSTREAEFEETNDNGVLVLPEAPKSHGVRGECGLDIDTMS	330
Db	272 gngnpppeefllylpgqpegirtsntyltxd-----vrrnatgdykcsllidkksmla	323
Qy	331 LUSEPOELLVNTVSDVRSVPAAP-ERQEGSSLTITCEASSODLEFQWLREFTGVLEBG	389
Db	324 stla---ltvhy1-dlslnpsgevtqlgdaipvscctlsasrnatvymkdnlr--lrss	376
Qy	390 PVLQLHDLKREAGGGRCAVSVPISPGINRLOLVNAVILFGPPWMAFKERKVAKKAMVPLN	449
Db	377 p---sfslslyhqdagnyvcetaldveveglkkrtesllilveqkp-qikmtkrtkdpjsikt	432
Qy	450 LSCSEASGHPRTISMVNVGTAS--EODQDPQ--RVLTSLVWLVTPELLTGVECTASN	503
Db	433 lichevqfbrpalaqwtlqsgsvlnqteespyingryys--kllspsent-ltctaen	489
Qy	504 DKGKNSILFLELVNLTITPDSNNTTGLSTASTASPHRANSTSEKRLPEPESKGVIV	563
Db	490 qterfvnslnvsaalsi---pe-----headelsdenr-ekvndgaktllv	530
Qy	564 AVIVGILVAVLAVLAVLYFLY-KKGRLLPCRRSGKQEIITPPSRKSE	607
Db	531 glivvglllaalvagvyvylwmkkskstaakhnkdilgmeeenkkle	575
RESULT	6	
ID	R97231 standard; Protein; 570 AA.	
XX	R97231;	
AC	R97231;	
DT	08-OCT-1996 (first entry)	
XX	Stem cell marker HCAPro.2.	
DE	Stem cell marker HCAPro.2.	
XX	Stem cell marker; haematopoietic cell antigen; HCA; HCASeq.2;	
KW	HCAPro.2.	
OS	Homo sapiens.	
XX	EP716146-A2.	
PN	12-JUN-1996.	
PD	02-FEB-1995; 95SEP-0300661.	
PE	06-DEC-1994; 94US-0352323.	
XX	(SYST-) SYSTEMIX INC.	
XX		

PI	Gearing D , Uchida N , Yang Z ;
XX	
DR	WPI : 1996-269990/28 .
DR	N-PDB ; T28820 .
XX	
XX	DNA encoding human haematopoietic cell antigen proteins - useful as
PT	stem cell marker proteins in functional studies and for antibody
PT	prodn.
PS	Claim 8; Fig 4B; 17pp: English.
XX	
CC	HCAPro.1 (#97230) and HCAPro.2 (#97231) are novel human
CC	haematopoietic cell antigens associated with stem cells.. They
CC	are the respective products of nucleic acid sequences HCASeq.1
CC	(T28819) and HCASeq.2 (T28820) obtd. from human CD34+ bone marrow.
CC	Expression systems are provided for prodn. of recombinant HCA
CC	proteins. The proteins are useful as stem cell marker proteins in
CC	functional studies, and can also be used to produce antibodies
CC	that allow the purification of stem cells from haematopoietic and
CC	other sources .
XX	
SQ	Sequence      570 AA;



QY 264 EGDREIIRC---LADGNPPPHFSISKONPSTREAEEETN-----DNGVILVEPARKHS 315  
 Db 1932 egdtltltcpiklaediadqvmvswtkdsraldgltdhvdidsdqgrkltlisqaalen 1991  
 QY 316 GRECCGLD-----LDTMISLSEPOELLVNVSDVRSPAPEROEGSSSLFTCEAES 369  
 Db 1992 gilytclalnraageaslefkveislsp-----vldistrndvppqavavngptlmrcavtg 2044  
 QY 370 SODLEFQWLR-----EETGVLERGPVYLQHLDLKREAGRCVAVSPISFGLNFTQL 422  
 Db 2045 hpfpslkwlnkgvevdenniriveggvqltrtdsdhagkwscvae--ndagvkelcm 2102  
 QY 423 VNVATGPPPMMAFERKRWVKKENMVLNLSCEASGHPRTISMNVNGTASQDDDPQRYLS 482  
 Db 2103 v-ldvftppvsvskdnpikalgeltlfcnaagppqglkwaaggslldspdgari 2161  
 QY 483 TLNVLVTPLELLETV---ECTASNDLGKNTSILFLELVNLTTLTPDSNTTGTSTSPAS 539  
 Db 2162 kgarldipnlkktcdvqdytcqalnaagttsea-----svsvdvlvpeinrdgid---msp 2213  
 QY 540 HFRANSTSTERKL---PEPESR 558  
 Db 2214 rlpagqsltlqclagqkpyppqr 2236  
 RESULT 8  
 Y08402  
 ID Y08402 standard; Protein; 1380 AA.  
 AC Y08402;  
 DT 24-JUL-1999 (first entry)  
 DE Drosophila sp. ROBO2 extracellular domain protein.  
 KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;  
 cell morphology; screening assay.  
 OS Drosophila sp.  
 PN MO9920764-A1.  
 PD 29-APR-1999.  
 PF 20-OCT-1998; 98WO-US22164.  
 PR 14-NOV-1997; 97US-0971172.  
 PR 20-OCT-1997; 97US-0062921.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodman CS, Kidd T, Mitchell KJ, Tear G;  
 DR WPI: 1999-312615/26.  
 DR N-PSDB: X57251.  
 PT Robo polypeptides, a new immunoglobulin superfamily member  
 PS Claim 1; Page 52-56; 80pp; English.  
 XX This invention describes novel Robo (roundabout) polypeptides, involved  
 CC in nerve guidance which have been isolated from Drosophila sp.,  
 CC C. elegans, human and murine samples. The products of the invention can  
 CC be used to raise anti-Robo antibodies, which can be used to modulate cell  
 CC function or morphology. The Robo polynucleotides and fragments are useful  
 CC as probes and primers and for production of the Robo polypeptides. The  
 CC probes and primers are also useful in screening assays.  
 SO Sequence 1380 AA;

Best Local Similarity 23.8%; Pred. No. 2.6e-09;  
 Matches 119; Conservative 72; Mismatches 193; Indels 116; Gaps 27;  
 QY 137 EEPNIOVNPIGIVNKSKEPEVATCYGRNGYPLPVOIWMYNGRPLLEEKRRVHIQSSQY 196  
 Db 2 enprlilehnpdltvpxkndp-ftfnca-egnpelptlqikdgrfkltdgshim-----1 55  
 QY 197 ESSGILYTLQSLILKAOLVKEEDKDAQEYCELVNRLPSGNHMKESREVVPPVYPREKWLVEV 256  
 Db 56 paqglfllkvlhr-----resdagtlwceaknef-----gvartsnalqvavilde--fll 106  
 QY 257 EPVG-MKEGDREIIRCILAD-GNPPPHFSISKONPSTREAEEETN-----DNGV 304  
 Db 107 epantlvaggevalmeagaprgsppqiswrtkg-----qtlolvgnkrlirivdggn 158  
 QY 305 LVLEPARKHSGRECCGLDLMISLSEPOELLVNVSDVRSP---AAPEROE---G 358  
 Db 159 latqeargsddgygc-----vknvngltrsataflk-vhvtrpflirgpnqatavga. 210  
 QY 359 SSLTINCE--AESQDL-----EFQWLREETGOV-LERGPVYLQHLDLKREA 401  
 Db 211 ssavfqcrriggdipdvltwrtasgmpirklfswlnsaagrvhvlcdsrklkiddvrlcd 270  
 QY 402 GCGYRCVAVSPISFGLNRTOLVNVATGPPPMMAFERKRWVKKENMVLNLSCEASGHPRT 461  
 Db 271 mgeytcead-navggitatgltv--happkfvrpknqlveigdevlfecqanghprpt 327  
 QY 462 ISMNVNGTAS-----EQDQDPQRYLSLTLNVLVTPLELET--GVECTASNDLGKNTSILF 515  
 Db 328 lywsevegasslllpgyrd-----grmevlllpegrsvlsiarfaredsgk----- 372  
 QY 516 LVNLTTLTPDSNTTGTSTSPASPHTRANSTSTERKLPEP-ESRGVIVAVIYICILVAV 574  
 Db 373 vvtcnal-----navgsvsrtlv-----svdtcfelpppliegpnvntlpvksivv-- 420  
 QY 575 LGAVLYFLYKKRKLPCRSG 594  
 Db 421 -----lpcrtlg 427  
 RESULT 9  
 Y13564  
 ID Y13564 standard; Protein; 1381 AA.  
 AC Y13564;  
 DT 30-JUL-1999 (first entry)  
 DE Drosophila Robo 2 polypeptide.  
 KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;  
 modulation; nerve cell function.  
 OS Drosophila sp.  
 PN MO9925833-A1.  
 PD 27-MAY-1999.  
 PF 13-NOV-1998; 98WO-US24327.  
 PR 14-NOV-1997; 97US-0065543.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Goodman C, Kid T, Mitchell KJ, Russell C, Tear G;  
 DR WPI: 1999-338008/28.  
 DR N-PSDB: X55768.  
 PT Modulation of Robo-Comm polypeptide interactions  
 PS Disclosure: Page 34-38; 56pp; English.

XX The invention relates to a method for modulating the amount of Comm  
CC (comm:sunreless) polypeptide in contact with a cell expressing active  
CC Robo (roundabout) on its surface. The method comprises modulating the  
CC effective amount of Comm polypeptide in contact with the cell, where the  
CC amount of expressed active Robo is specifically modulated inversely with  
CC the modulation of the effective amount of Comm in contact with the cell.  
CC The method is used to modulate the amount of active Robo expressed on a  
CC cell. The method can be used to screen for agents that modulate Robo:Comm  
CC interactions. This is particularly useful for modulating nerve cell  
function.

50 Sequence 1381 AA;

Query Match	7.5%;	Score 252;	DB 20;	Length 1381;
Best local similarity	32.8%;	Score 343;	DB 20;	Length 1381;

```

QY      137 EEPBIOVNPGLGIPVNSKEPEEVAIVCGRNGYDIPQVIWYKNGRPJLKEKNVHIQSSQTV 196
Db      2   enpclihepmcltvpknd- ftfnqca -egnpctpiqyfdkgrelktgshrim----- 1
QY      197 ESSSLRYTLOSILTAOLVTKEDKOAEFCELNRYRLPSGNHMKESREVTVPVFPTFKVLEV 256
Db      56 pagglfllfklythr-----resdaglyceaknef-----gvassnatclqvavlide--frl 106
QY      257 EPVG-MKEEDRVLRCLAD-GNPPHFESISKONBSTREAEEETTN-----DNGV 304
Db      107 epantlvaggevalmeccaprgspdepqjswkrng-----qlnlvgnrkrlivdggn 158
QY      305 LVLEPARKREHSRGFEQGLDTPMISLSEPEPELLNVSDVRSP--AAPEHQE--G 358
Db      159 lajqearqsddgyqc-----vkknvygttestatfik-vhvrflligpqqaav 210
QY      359 SSLUTICE-AESSQDL-----EFQWLREETGOV-LERGPVLQHLDKREA 401
Db      211 ssavfgcrtlgdgdldpvdwlrtaasgmplrkfwslhaasgrhvlnedrlkliddvled 270
QY      402 GGGYRCASVSPISPGINRTQLNVALIEGPMMAFPERKVWVKENMVNLSCSAAGHPRP 461
Db      271 mgeyltcoad-navgyltalqtlvt-hpprkfvirpkqlveiydevlfeeqdanghprc 327
QY      462 ISMNVMGTAS-----EQODDPQRVLSTLNALVTPBELLET-GVECTASNDLGKNTSILFLE 515
Db      328 lyvsvegnesallpyrd-----grmvevlltpgezvsiaarfaredesgk----- 372
QY      516 LVMLTLTLPDSMTTGILSTASTASPHTRANSSTEKULPEE-ESRGVVIVAVIICUILLAV 574
Db      373 vvecnml-----navgsvsarltvv-----svdtqtelrpppliieggpvngqlipvksltvv--- 420
QY      575 LGAVLVFLYKKKGKLPCRMSG 594
Db      421 -----lportlg 427

```

```

RESULT 10
R39686
ID R39686 standard; Protein; 647 AA

```

AC R39686;

DT 23-DEC-1993 (first entry)

VCAM-6D/ICAM4-1.

KM vascular cell adhesion molecule; monoclonal antibody; chimera: 4B9  
 KM VLA-4; expressing cells; intracellular adhesion molecule; VCAM-7D;  
 KM VCAM-6D; substitution; Ig superfamily; homology.  
 KM

OS Synthetic.

XX

Key	Location/Qualifiers
FH	1..24
FT	/note="Signal peptide"
FT	25..647
FT	/note="Mature protein"

PN W09314220-A.

PD 22-JUL-1993.

PF 12-JAN-1993; 93WO-US00031.  
YY

PR 13-JAN-1992; 9205-0821712.  
XX  
XX

PA (BLOJ ) BIOGEN INC.  
XX

XX  
XX  
XX

Benjamin C., Esq.

DR N-PSDB; Q46662.

PT Monoclonal anti

PT leukocyte binding to endothelium e.g. post-reperfusion injury, etc.

PS Disclosure; Page 83-86; 108pp; English.

CC virus sequence represents a chimeric vascular cell adhesion molecule  
 CC (VCAM)/ICAM polypeptide which was used to determine regions of VCAM-1  
 CC involved in recognition of VLA-4-expressing cells. Recombinant genes  
 CC encoding VCAM/ICAM (intracellular adhesion molecule) chimeras were  
 CC produced by existing portions of the VCAM-7D and VCAM-6D cDNAs and  
 CC replacing them with analogous regions of ICAM-1 using unique  
 CC restriction endonuclease sites within the VCAM-1 cDNA, and PCR  
 CC generated fragments of ICAM-1 (see also Q43578-84). Substitution  
 CC of the excised VCAM-1 regions was done to eliminate as much as  
 CC possible structural distortions that would accompany deletion of one  
 CC or more domains of the VCAM-1 molecule. ICAM-1 was chosen as a donor  
 CC because it is also a member of the Ig superfamily and has the highest  
 CC degree of amino acid homology to VCAM-1 among superfamily members.  
 CC The N-terminal three domain of VCAM-1 support adhesion to VLA-4, and  
 CC VCAM/ICAM-1 and VCAM/ICAM-2 in which most of the first domain has been  
 CC replaced by ICAM-1, do not bind to the anti-VCAM-1 monoclonal antibody  
 CC (MAb), 4B9. Therefore the epitope which binds to 4B9 must be situated  
 CC in domain 1.

Sequence 647 AA;

Query Match	7.3%;	Score 244.5;	DB 14;	Length 647;
Best Local Similarity	21.5%;	Pred. No. 3e-09;		
Matches 147;	Conservative 100;	Mismatches 255;	Indels 181;	Gaps 32

```
QY   31 PAELVEVEVSTALIKCGLSSQGNSHVDWFVSVAHEKRLIFRVROGQGGSEPEYEQ    90
      | : :: : | :          ||| |: ::
Db   31 pgpr-iaeqidsvmltcsvmgces--psfsw-----rtqidspsjgkvise-----    74
```

91 RLSLDRGATLALTQVTPQDERIFLCQ--GKRPRSQEYRIQLRVYKAPPEPNIQVN-PL 146

DB /5 -----gtnstltlspvstenehsylctvtcgh--kklekqivelysfprkdpelhsqpl 127

14/ --GLPVNSKEPEVATCVGRNGYPIQV-IWKNGRPLKEEKNRVHIQSSQTVESSGL-Y 202

```

120 eaqkprlvk-----csvadvypidrleldllkgdhlmsqetledadrksletkslev 180

```

[illegible]

caignvvlciainuamusvprvlgavkeiqvylsprkrlvlsvlhpsr 234

[illegible][illegible]

```

OY 321 OGDLDTMTLSLSEPELLVNVSDVRSVPAAPERQ-----EGSSLTLTCEAES---SQ 371
DB 292 egvnl---lgnrkveell-----vgafprdpelmsgglvngssvsvckvpsvypld 342
OY 372 DLEFQWLREETGOVLERGFPVLOLHDLK-----REAGGRCVAVS----- 411
DB 343 rlelletlkgct--lleniefledtmkslenkslemftlptledtgkalcvcgklnhdmm 400
OY 412 ---PSTPGINRQOLVNVAFGPPMAFKERKVVAKENMVL-----NLSCASGHRPTI 462
DB 401 efepkqrgstqcllyvna---p---rdttvlyvpspslleegssvnmctlsqgfapkl 452
OY 463 SNMNVNGTASEODDPORVLTSLNVLTPPELLETGVE-----CTASNDGKNTSILFEL 516
DB 453 lvs-----rqlpngelqplsenatliltstkmcdsgvylceglngqgrsrkveell 504
OY 517 -----VNLTTLTLPDS-----NTTGLSTSTASPHT-----RANSTSTERK 551
DB 505 gvtpkdkitafpsesvkegdvltscgnypetwlllkkkaetgdtvlskldgagyltr 564
OY 552 -----LPEERSRGV-----IVAVICILVAVLGAV 578
DB 565 kaqlkdaqvcecskxkvsgqlrslltdvgrenkdyfispellvlyfssallipalqm 624
OY 579 LVFLYK---KGKLPGRSGKOEI 598
DB 625 lyfarkamkgysylveagkskv 647

RESULT 11
R39682
ID R39682 standard; Protein: 736 AA.
XX
AC R39682;
XX
DE 23-DEC-1993 (first entry)
XX
DE VCAM/ICAM-2.
XX
KW Polymerase chain reaction; primer; PCR; amplify; VCAM-1; ICAM1;
KW vascular cell adhesion molecule; monoclonal antibody; chimera; 4B9;
KW VLA-4; expressing cells; intracellular adhesion molecule; VCAM-7D;
KW VCAM-6D; substitution; Ig superfamily; homology.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..24
FT Protein /note="Signal peptide"
FT Protein /note="Mature protein"
XX
XX WO9314220-A.
XX
XX 22-JUL-1993.
XX
XX 12-JAN-1993; 93WO-US00031.
XX
XX 13-JAN-1992; 92US-0821712.
XX
XX (BIOJ ) BIOGEN INC.
XX
XX Benjamin CD, Osborn L;
XX
XX WPI; 1993-243235/30.
XX
XX N-PSDB; Q43588.
XX
XX Monoclonal antibody to epitope on the fourth Ig-like domain of
XX VCAM-7D - for treating inflammation or disease associated with
XX leukocyte binding to endothelium e.g. post-reperfusion injury,
XX etc.

```

```

PS Disclosure; Page 56-59; 108bp; English.
XX
CC This sequence represents a chimeric vascular cell adhesion molecule
CC (VCAM)/ICAM polypeptide which was used to determine regions of VCAM-1
CC involved in recognition of VLA-4-expressing cells. Recombinant genes
CC encoding VCAM/ICAM (intracellular adhesion molecule) chimeras were
CC produced by existing portions of the VCAM-1 molecule and VCAM-6D cDNAs and
CC replacing them with analogous regions of ICAM-1 using unique
CC restriction endonuclease sites within the VCAM-1 cDNA, and PCR
CC generated fragments of ICAM-1 (see also Q43578-84). Substitution
CC of the excised VCAM-1 regions was done to eliminate as much as
CC possible structural distortions that would accompany deletion of one
CC or more domains of the VCAM-1 molecule. ICAM-1 was chosen as a donor
CC because it is also a member of the Ig superfamily and has the highest
CC degree of amino acid homology to VCAM-1 among superfamily members.
CC The N-terminal three domain of VCAM-1 support adhesion to VLA-4, and
CC VCAM/ICAM-1 and VCAM/ICAM-2 in which most of the first domain has been
CC replaced by ICAM-1, do not bind to the anti-VCAM-1 monoclonal antibody
CC (Mab), 4B9. Therefore the epitope which binds to 4B9 must be situated
CC in domain 1.
XX
SQ Sequence 736 AA:
OY
DB 33 PELVEV-----EVGSTALLKCGLSQSGNLSHVDFSVHKEKRTLIFRVGGQSGSEP 85
DB 112 pervalaprpwpgvgnklltrc---gveggapran-----ltvlllrgeklmksgef 161
OY 86 GEYEQRLSDRGATLALQVTPODERIFLCQKR-----PRQERYTOLRVYKAP 137
DB 162 ledadrtksletkslevlfrpvedlqvlvcraklnhdemsdvtrvgavkqlvayispk 221
OY 138 EPNIOVNPGLIPVNSKEPEEVATCVGRNGYPIQOVLIWK---NGRPLKEKNVHIOSSQ 194
DB 222 ntvisvnp---sklqegsgvmtcscseqlpapeitfsskklidng-----nlqhlsgna 271
OY 195 T-----VSSGLYTLQSLTILKAOLYKEDKDAOFYCELYNLRPSGNHMKESREVPVVFY 247
DB 272 tlliammedsglylcycegy---nll-----gnkrke-----veliv 305
OY 248 PTEKWLVEVPVGM---KEDDRVEIRCLADGNPPPHSISKQNPSTREAEETTNDNGVL 305
DB 306 qekpfveaisppriaqldgsvmltcsymgscpsfswtqldspisgkvrsqgnstsl 365
OY 306 VLEPARKHSGRYEC-----OGDLDTMTLSLSEPELLVNVSDVRSVPAAPERQ--- 356
DB 366 tispvsfenehsylctvtgchkklekgyivelysfidp-----pelmsg 410
OY 357 ---EGSSLTLTCEAES---SQDLEFQWLREETGOVLERGFPVLOLHDLK----- 398
DB 411 glvngssvsvckvpsvypldrlelletlkgct--lleniefledtmkslenkslemftl 468
OY 399 ---REAGGRCVAVS-----PSTPGINRQOLVNVAFGPPMAFKERKVVAKENMV 447
DB 469 priedtgkalcvcgklnhdmmefepkqrgstqcllyvna---p---rdttvlyvpspsl 550
OY 448 L-----NLSCASGHRPTISNMNVNGTASEODDPORVLTSLNVLTPPELLETGVE--- 498
DB 521 leegssvnmctlsqgfapklivs-----rqlpngelqplsenatliltstkmcdsg 572
OY 499 ---CTASNDGKNTSILFEL-----VNLTTLTLPDS-----NTTGLSTSTASPHT--- 541
DB 573 vylceglngqgrsrkveelllygtpkdkitafpsesvkegdvltscgnypetwlll 632
OY 542 ---RANSTSTERK-----LPEERSRGV----- 561
DB 633 kkaetdvtlksldgagyltrkaqlkdaqvcecskxkvsgqlrslltdvgrenkdyf 692
OY 562 ---IVAVICILVAVLGAVLVFLYK---KGKLPGRSGKOEI 598

```

Db 693 spellvlyfssallipaligmllyfarKamkysylveadkskv 736

RESULT 12

ID Y53666 standard; Protein: 4412 AA.

AC Y53666;

XX 22-FEB-2000 (first entry)

DE Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.

XX Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;

KM bone development; g1/1017427/emb/CAA62189.

XX Unidentified.

OS

XX WO960164-A1.

PN

XX 25-NOV-1999.

PD

XX 14-MAY-1999; 99WO-US11066.

PF

XX 15-MAY-1998; 98US-0085673.

PR

XX (OUAR-) QUARK BIOTECH INC.

PA

PI Elnat P, Mor O, Skalter R, Feinstein E, Faerman A;

XX WPI; 2000-053304/04.

DR

XX Identification of stress induced genes for determining risk and

PT preventing, treating or controlling osteoporosis -

XX

XX Claim 32; Fig 6A-R; 308pp; English.

PS

XX The present sequence is obtained from a clustal X alignment with

CC protein 608. Protein 608 was identified using the method of the invention

CC after subjecting rat osteoblasts to mechanical stress. Expression of the

CC 608 gene was found to be upregulated by about 3-fold in cells subjected

CC to mechanical strain. The specification describes a method for the

CC identification of genes responsive to a specific mechanical stress. The

CC method comprises applying the mechanical stress to an organism (tissue

CC or cells comprising bone cells). Isolating the specific cellular

CC fractions and extracting mRNA from them, and differentially analysing the

CC mRNA in comparison with control samples. The method is used to identify

CC genes whose expression is responsive to a specific stress. The identified

CC genes are employed in determining risk associated with a physiological or

CC disease state. The risk determination methods are used for testing a

CC medication for gene therapy. These medications, or genes identified by

CC the method of the invention, are used for treating, preventing or

CC controlling a physiological or disease state (especially osteoporosis or

CC bone density or other factors causing or contributing to osteoporosis or

CC its symptoms or other conditions involved in mechanical stress or its

CC lack. The methods can also be used for advancing research or studies in

CC bone development.

XX

SQ Sequence 4412 AA:

Query Match 7.0%; Score 236; DB 21; Length 4412;

Best Local Similarity 21.6%; Pred. No. 1.8e-07;

Matches 117; Conservative 88; Mismatches 231; Indels 106; Gaps 24;

QY 34 ELVEVEVSTALLKCGISQSGNLSHVDWFSVHKERKTLIFRVQGGQSGSEPEYQRLS 93

Db 1000 epleaavgsdvslyqcyagvpe--llvswy-----kgdktlrpp-eyrly 1042

QY 94 LQDRGATLALTYTPROBERFLCO-----GKRPRQSEYRQLRVK---APEEPNQQVP 145

Db 1043 flnnvalivlnkynlndsgyctkaenslytasskvtvrlqerqlpdpstfargldie-qt 1101

QY 146 LGIPVNSKEPEEVAACVGNHGYPIPOVIWYKNGRLKEKNRHHIOSSQTFESSGLYTLQ 205

Db 1102 vglpv-----tlcrlngsapi-qvcwyrdgvllrhen---lqsfvnvalclklly 1150

QY 206 SILKAOLVKEKDQAFYCELANTRLDPSGNHMKESREVTVPVPEYKWLVEVPGM-LKE 264

Db 1151 tdl-----shsgqyscsasn--plglaassar--ltarepkkspfldikpsidvia 1198

QY 265 GDRVEIRCLADGNPPPHFISIKONPSTREAEETW-----NDNGVLVLPARKHSGREY 320

Db 1199 gesadfechvgaqpmrltwsdkmkeirpganytlrcvgnlphrltk-vgkgdsgytc 1257

QY 321 Q-----GLDLDTMISLSEPOLLVNVYSDVRSAPABERQSSLTLTCEASSQDLEF 375

Db 1258 gqndvkgkmcasqslsvkepk-----fkkleaskva---kgesigtleckisgspelkv 1310

QY 376 QMLREETG-----QVLERGPVLQHLDLKREAGGGRCAVSPSTIPGINRQL-VNV 425

Db 1311 swfrndselheswkymfsfinsvallltlnaasadsdyiceah---ngvgdascstal 1366

QY 426 AIFGPPWMAFKRKRYWVENMYLNLSCENSGHPRFTISMVNVNGTASBEDQDPQRYLSTLN 485

Db 1367 tvkappvftqkpsvgaalkgsdvlqceisgtfpfevw-----vkdrkqyrnsk 1417

QY 486 VLVTPELLETGY-----ECTASNDLGKNT---SILFELVNLTLTPDSWTT 529

Db 1418 flttskfrtdnlhlnleasdvgeyhckatnevgsdltscsvkfkepprfvkkisdctsl 1477

QY 530 TG 531

Db 1478 ig 1479

RESULT 13

ID Y88565 standard; Protein: 848 AA.

XX Y88565;

AC Y88565;

XX 07-AUG-2000 (first entry)

DT

XX Human NCAM 140kd isoform precursor amino acid sequence.

DE

XX NCAM; neural cell adhesion molecule; Ig1; immunoglobulin domain 1;

KW neurite outgrowth promoter; proliferation; immune damage; sclerosis;

KW impaired myelination; stroke; Parkinson's disease; memory; schizophrenia;

KW Alzheimer's disease; diabetes mellitus; circadian clock; nephrosis;

XX treatment; prosthetic nerve guide; treatment; nervous system.

OS

XX Homo sapiens.

PN WO200018801-A2.

XX

XX 06-APR-2000.

PD

XX 23-SEP-1999; 99WO-DK00500.

PF

XX 29-SEP-1998; 98DK-0001232.

PR 29-APR-1999; 99DK-0000592.

XX

XX (RONN/) RONN L C B.

PA (BOCK/) BOCK E.

PA (HOLM/) HOLM A.

PA (OLSE/) OLSEN M.

PA (OLSE/) OLSEN M.

PA (OSTE/) OSTERGAARD S.

PA (JENS/) JENSEN P H.

PA (POUL/) POULSEN F M.

PA (SORO/) SOROKA V.

PA (RALE/) RALETS I.

PA (BERE/) BEREZIN V.

PI Ronn LCB, Bock E, Holm A, Olsen M, OSTERGAARD S, Jensen PH;

PI Poulsen FM, Soroka V, Ralets I, Berezin V;  
 XX WPI; 2000-293111/25.  
 XX  
 PT Compositions that bind neural cell adhesion molecules useful for  
 PT treating disorders of the nervous system and muscles e.g. Alzheimer's  
 XX and Parkinson's diseases -  
 PS Disclosure: Fig 17; 119pp; English.  
 XX  
 CC This sequence represents the human neural cell adhesion molecule (NCAM)  
 CC amino acid sequence. NCAM is found in three forms, two of which are  
 CC transmembrane forms, while the third is attached via a lipid anchor to  
 CC the cell membrane. All three NCAM forms have an extracellular structure  
 CC consisting of five immunoglobulin domains (Ig domains). The Ig domains are  
 CC numbered 1 to 5 from the N-terminal. The invention relates to a compound  
 CC containing a peptide which binds to the NCAM Ig1 domain. The compound  
 CC binds to NCAM-Ig1/Ig2 domains, and is capable of stimulating or promoting  
 CC neurite outgrowth from NCAM presenting cells, and is also capable of  
 CC promoting the proliferation of NCAM presenting cells. The compound may be  
 CC used in the treatment of normal, degenerated or damaged NCAM presenting  
 CC cells. The compound may in particular be used to treat diseases of the  
 CC central and peripheral nervous systems such as post operative nerve  
 CC damage, traumatic nerve damage, impaired myelination of nerve fibres,  
 CC conditions resulting from a stroke, Parkinson's disease, Alzheimer's  
 CC disease, dementia, sclerosis, nerve degeneration associated with  
 CC diabetes mellitus, disorders affecting the circadian clock or  
 CC neuro-muscular transmission and schizophrenia. Conditions affecting the  
 CC muscles may also be treated with the compound, such as conditions  
 CC associated with impaired function of neuromuscular connections  
 CC (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).  
 CC Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and  
 CC II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated  
 CC using the compound. The compound is used in a prosthetic nerve guide, and  
 CC also to stimulate the ability to learn, and to stimulate the memory of a  
 CC subject.  
 CC  
 XX  
 SQ Sequence 848 AA:  
 7.0%; Score 234; DB 21; Length 848;  
 Best Local Similarity 21.1%; Pred. No. 2.4e-08;  
 Matches 126; Conservative 86; Mismatches 222; Indels 164; Gaps 31;  
 Query Match  
 33 PELVEVVGSTALLKGLSGSGNLHVD--WFSVHEKRTILFRVGGGSEPEEYQ 90  
 26 psgqelsvgsckfflc---qvagdakdkdswfsgngklt-----pnqg 67  
 91 RLSTL---QDRGATLALTOVTPQDERIFLC--QGRKRRSQEYRQLRVYKAPERNQVNP 145  
 68 rlsvvvndssstltltnanidaglykcvlvgdgseseatvnnkifg----kimfn 122  
 146 LGIPVNSKEEP--VATCVGRNGYRIPQVITWKNRPLKEKN--RVIIQSSQVSESSGLT 203  
 123 apfpgfregedavivcdvssip-pltllkhkgivdlkkdrifvlssnnylqitg--- 178  
 204 LGSILKAQLVKEKDKDOFCELVNLRPSGNHMKESREYVVPVFPPEKYALEVEPYGMUK 263  
 179 -----ikktdeglytce-----grllarge-----infkdiqylvnnppliqgr 217  
 264 E-----GDRVEIRCLADGNPPHFSISKONPSTRAEET-----INDNGVILDEPA 310  
 218 gnlvntanlqsvlvcdaegfpeptmwtckdgdqiegedekylfsdssqlltkkv 277  
 311 KRHSRRYECG-----LDLTMTLSLSEPELLVNVSDVSPAPEROEGSSLTLT 364  
 278 dkndeaeiyiclaenkagqdatlhlkvfakp---itye-----nqtmeltee--gvlt 328  
 365 CEAVESODLEFQW-----LREET--GQVLERGPV-----LQHLDKREAGGRCVCA 410  
 329 ceasgdpipsticwtstnmsiseekldgmvmvsharvstllkslygtdeageyictas 388  
 411 -----VPSIPGLNRTOLVNVAFISPPWMAFKERKRVWVKENNVLNLSCA 454

DB 389 ntiqgdsgsmylexyqapkiqg-----pvavy-----tw---egngvnlitcey 428  
 QY 455 SGHPRPTISWNVGTASEQDQDQVRLSTLNVVTP--ELLEET-----GVECTASN 503  
 DB 429 faypsatlswfirgqll-----psnysnkiynpcsaalyevtpdsendfynmctavn 483  
 QY 504 DLGNKTSILFLLELVNLTTLTPDSNTTGTGSTASPHSTRANSTTERKLPESPESRGV 561  
 DB 484 riqges-----lefilvgadtspss-----slgdvepy-----sstaqvgtdeeatgyv 528  
 RESULT 14  
 R08117  
 ID R08117 standard; protein; 647 AA.  
 XX  
 AC R08117;  
 XX  
 DT 27-FEB-1991 (first entry)  
 XX  
 DE Vascular cell adhesion molecule 1 from pCDM8 clone 41.  
 XX  
 KW Endothelial cell-leucocyte adhesion molecule 1; ELAM1;  
 KW vascular cell adhesion molecule; VCAM1; antibodies;  
 KW molecule involved in leucocyte adhesion; M1A; inflammation.  
 XX  
 OS Homo sapiens.  
 PN W09013300-A.  
 PD 15-NOV-1990.  
 XX  
 PF 27-APR-1990; 90WO-0502357.  
 XX  
 PR 18-DEC-1989; 89US-0452675.  
 PR 28-APR-1989; 89US-0345151.  
 PR 01-JUN-1989; 89US-0359516.  
 XX  
 PA (BIOG-) BIOGEN INC.  
 XX  
 PI Hession C, Lobb RR, Goeltz SE, Born L, Benjamin CD;  
 PI Rosa MD;  
 DR N-PSDB; Q06687.  
 DR WPI; 1990-361248/48.  
 PT Endothelial cell adhesion mols. - M1As and DNA encoding them and  
 PT inhibition-detection of binding of leukocytes to endothelial  
 PT cells  
 PS  
 XX  
 CC Disclosure: Fig 3 (A-D); 136pp; English.  
 CC  
 CC The VCAM1 polypeptide possesses a hydrophobic N-terminal  
 CC characteristic of a signal sequence. It is predicted that the  
 CC N-terminal amino acid of the mature protein will be phenylalanine 25.  
 CC The extracellular domain of the polypeptide is ca. 606 amino acids  
 CC including the signal sequence and is followed by a hydrophobic  
 CC transmembrane region of 22 amino acids. The protein possesses a  
 CC short charged cytoplasmic tail of 19 amino acids. The protein  
 CC contains six potential N-glycosylation sites.  
 CC Comparison of VCAM1 and VCAM1b (Q06688) revealed that they are  
 CC virtually identical except for one significant difference: VCAM1b  
 CC contains an insertion of 276 nucleotides near the middle of the  
 CC coding region. These nucleotides encode 92 additional amino acids  
 CC which form an extra domain of 84 amino acids situated between the  
 CC end of VCAM1 domain 3 and the beginning of VCAM1 domain 4. This  
 CC domain is designated domain 3B.  
 CC Cells expressing ELAM can be used to identify mols. which inhibit  
 CC binding of leucocytes to endothelial cells, and such binding can  
 CC be inhibited (i.e. inflammation can be treated) using an ELAM  
 CC (or fragment), antibodies which recognise M1A, ELAM ligands or  
 CC their fragments, carbohydrates which bind to ELAM and antibodies  
 CC which recognise ELAM.



```

QY 382 TGOVLERGVLQDLHDK-----REAGGGRVAVSY-----PSIPGLN 418
Db 353 t--lleenlefedtdmkslenkslemftipiedtgkaihvqaklhiddmetepkqrst 410
QY 419 RTQLVNVAlFSPPMMAFKERKVMVKNVNL-----NLSCASGHPRTISMNVNGTASE 472
Db 411 qllyvva---p---fdtvlvpspsilleegssvnmctlsqgfapakiws----- 455
QY 473 QDQDPQVRLSTLNLVLPPELLETGVE-----CTASNDLGKNTSILFEL-----VNLTF 520
Db 456 -rqlpngelqpsenatltlistkmedsgvylceqingagrstrkevelilqvtpkdikt 514
QY 521 TLTPTS-----NTTGLSTSTASPT-----RANSTSTERK-----LP 553
Db 515 afpsesevkegdvtliscctcgnvpelwllkkkaetgdtvlksidgaytirkaqlkdagvy 574
QY 554 EPESRGVY-----IVAVIVCTILVAVLGAVLYFLYK---K 585
Db 575 ecesknkvysqlrsltdvggrennkdyfspellivlyfasslilpaigmilyfarknmk 634
QY 586 GKLPORRSGKQEI 598
Db 635 gsyslveaqkskv 647

```

Search completed: March 23, 2001, 12:31:40  
 Job time: 63 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:37 ; Search time 22.65 Seconds  
(without alignments)  
1936.593 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGPRVCAFLAACCCCPR.....SSGDKRAPGDGKEYIDLRH 646

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3327	98.9	646	2	cell surface glyco
2	863	25.7	584	2	s-glycerin precurs
3	647.5	19.3	628	2	Lutheran blood gro
4	606	18.0	588	2	B-CAM protein - hu
5	509.5	15.2	588	2	adhesion molecule
6	505.5	15.0	588	2	surface glycoprote
7	502.5	14.9	587	2	DM-GRASP precursor
8	492.5	14.6	583	2	alcam - human
9	457	13.6	523	2	neurotin - goldf
10	281	8.4	5175	2	hypothetical prote
11	281	8.4	5198	2	hemocentin precurs
12	270	8.0	764	2	irregular chlam C
13	264.5	7.9	538	2	kinase-like protei
14	259.5	7.7	1051	2	neural cell adhesi
15	256	7.6	1051	2	neural cell adhesi
16	251.5	7.5	4391	1	perlecan precursor
17	247	7.3	1612	2	ductal protein - mo
18	246.5	7.3	725	1	neural cell adhesi
19	246.5	7.3	739	2	neural cell adhesi
20	246.5	7.3	1651	2	transmembrane rece
21	244.5	7.3	858	1	neural cell adhesi
22	244.5	7.3	3707	2	heparan sulfate pr
23	242.5	7.2	853	1	neural cell adhesi
24	237	7.0	725	2	neural cell adhesi
25	237	7.0	1092	1	neural cell adhesi
26	236	7.0	7962	1	neural cell adhesi
27	235	7.0	1091	1	neural cell adhesi
28	234	7.0	761	1	neural cell adhesi
29	233.5	6.9	1115	1	neural cell adhesi

## ALIGNMENTS

Result	1	138049	cell surface glycoprotein MUC18 precursor - human
N:Alternate names:	melanoma-associated glycoprotein MUC18 precursor		
C:Species:	Homo sapiens (man)		
C:Date:	01-Mar-1996 #sequence-revision 01-Mar-1996 #text-change 21-Jul-2000		
C:Accession:	138049; A34507		
R:Refs:	C. J. Kirsch, K. Rothbacher, U. J. Rietmuller, G. Johnson, J. P. Proc. Natl. Acad. Sci. U.S.A. 90, 8514-8518, 1993		
A:Title:	Genomic organization of the melanoma-associated glycoprotein MUC18: implicat		
A:Reference number:	138049; MUID:93391384		
A:Accession:	138049		
A:Status:	preliminary		
A:Molecule type:	DNA		
A:Residues:	1-646 <RES>		
A:Cross-references:	EMBL:X68264; NID:9433891; PIDN:CAA48332.1; PID:9825693		
R:Refman:	J.M.; Rietmuller, G.; Johnson, J.P. Proc. Natl. Acad. Sci. U.S.A. 86, 9891-9895, 1989		
A:Title:	MUC18, a marker of tumor progression in human melanoma, shows sequence simil		
A:Reference number:	A34507; MUID:90099368		
A:Accession:	A34507		
A:Molecule type:	RNA		
A:Residues:	1-71, 'SSVCARARANUGSTK', 91-120, 'LGPRSTASSASTKLM', 139-587, 'AAVALREAG		
A:Cross-references:	GB:M29277; GB:M28882		
A:Note:	this sequence has been corrected in 138049		
C:Genetics:			
A:Gene:	GDB:MCAM; MUC18; CD146		
A:Cross-references:	GDB:304548; OMIM:155735		
A:Introns:	23/1; 64/3; 134/1; 157/3; 187/1; 287/3; 342/1; 381/3; 429/1; 469/3;		
C:Keywords:	glycoprotein; transmembrane protein		
Query Match	98.9%; Score 3327; DB 2; Length 646;		
Best Local Similarity	98.9%; Pred. No. 2, 6e-205;		
Matches	639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;		
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		vascular cell adhe
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		neural cell adhesi
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		rig-1 protein - mo
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		vascular cell adhe
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		fasciclin II, tran
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		neural cell adhesi
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		platelet-endotheli
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		carcinoembryonic a
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		protein-tyrosine k
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		nephlin - human
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		B-cell adhesion pr
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		neural cell adhesi
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		advanced glycosyla
Qy	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		cell adhesion mole
Db	1 MGPRVCAFLAACCCCPRVAGVGEAEOPAPDELVEVEGSTALLKCGISGQNLSHV 60		

Oy 241 VTVPVFVPTKRVMLVEFVPGVMLKEGDRVEITRCLADGNPPPHSTSKONPSRRAEEETT 300  
 Db 241 VTVPVFVPTKRVMLVEFVPGVMLKEGDRVEITRCLADGNPPPHSTSKONPSRRAEEETT 300  
 Oy 301 DNGVLTVEPARKEHSGRVECOGLDPTMISLSEFOELLVNVSDVRYSPAAPEROGSS 360  
 Db 301 DNGVLTVEPARKEHSGRVECOGLDPTMISLSEFOELLVNVSDVRYSPAAPEROGSS 360  
 Oy 361 LTLTCEAASSODLEFQWMLRETFQVLEKRGVPLQHDCLKREAGGGRCAVASPTPGNRT 420  
 Db 361 LTLTCEAASSODLEFQWMLRETFQVLEKRGVPLQHDCLKREAGGGRCAVASPTPGNRT 420  
 Oy 421 QLVNVAIGPMPMAFKERKVVVKEMVNLSCSEASGHPRPTISNVNGTASEODODPORV 480  
 Db 421 QLVNVAIGPMPMAFKERKVVVKEMVNLSCSEASGHPRPTISNVNGTASEODODPORV 480  
 Oy 481 LSTLNLVLTPELLEFGVCTASNDLGKNTSLFLELVNLTLTDPDSNTTGLSTASPH 540  
 Db 481 LSTLNLVLTPELLEFGVCTASNDLGKNTSLFLELVNLTLTDPDSNTTGLSTASPH 540  
 Oy 541 TRANSTERKLPPEESRGVIVAVIYICILVAVLGAVLYELYKKGKLPCHRSKQEIITL 600  
 Db 541 TRANSTERKLPPEESRGVIVAVIYICILVAVLGAVLYELYKKGKLPCHRSKQEIITL 600  
 Oy 601 PPSKSELVVEKYSKDLPEEMGLLOGSSGDKRAPDOCEKTYIDLRH 646  
 Db 601 PPSKSELVVEKYSKDLPEEMGLLOGSSGDKRAPDOCEKTYIDLRH 646  
 RESULT 2  
 150419  
 s:glcerin precursor - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: 150419  
 R:Taira, E.; Takaha, N.; Taniguchi, H.; Kim, C.H.; Maki, N.  
 Neuron 12, 861-872, 1994  
 A>Title: Molecular cloning and functional expression of glcerin, a novel cell adhesion p  
 A:Reference number: 150419; MUID:94213753  
 A:Accession: 150419  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-584 <Tail>  
 A:Cross-references: GB:D38559; NID:g1009246; PIDN:BA07563.1; PID:g559701

```

Query Match      25.78; Score 863; DB 2: Length 584;
Best Local Similarity 35.18; Pred. No. 1e-47;
Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps

Qy      8 CAFLAACCCCPRYAGVGEAEPAPELVETVEVSTALTKCGL-----SOSQGNLSHVDW 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      22 CFFLLCLLCC-----GAAGR-----EYVMSSA-----GSGDRRHSGAGVQLLN-PW 63

Qy      63 FSVHKEKPTLIF--VRQSG-----OSEPEYDQSLSDRGATLALQVTPQDE 111
      | : : : : : : : : : : : : : : : : : : : : : : : :
Db      64 ECLTLRRVVLCOQRPRLTSGEAVNHNAGVVIDETEXSERLSVGEDRA-LISKVTRQDN 122

Qy      112 -RIEFC--GKRPRQEQRIQLRYKKAPEERINQVPLGIPVNSKEPEEATGACGRNGY 168
      | | | | | | | | | | | | | | | | | | | | | | | |
Db      123 APTFCQVQADSGQGVGSESTELYTKIPAPETIRPNSGIRPAQSDNDMLKTAQSTENSP 182

Qy      169 IPQVITYKNGRPLKEKNRNHIQSSQTVSESSGLTYLQSLKAQLVKEDKDAQFYCELN 228
      | : | | | | | | | | | | | | | | | | | | | | | |
Db      183 SPNITWYKNGEPLDQEEKTKILTLFVRESNGLTAVVSTLFSKTREDRNSLFFHCITWY 242

Qy      229 LPSGNHMKESREYVTPVRYPIPEKWKLEER-VGMLKGGDRREINCLADGNRPPIHFSKQ 287
      | : | | | | | | | | | | | | | | | | | | | | | :
Db      243 LOGQKRTKSDSPVNTVTEYPIEHELRAVTAAGLVKREDVDKLVCDADGNAPVSEFFRR 302

Qy      288 --NPSREAEETTDNDQNVLELPARKHSGRGCQGLDITMTLSLSEQDILVNVVSD 345
      | | | | | | | | | | | | | | | | | | | | | | | |

```

[illegible]

```

RESULT      3
138000
Lutheran blood group glycoprotein precursor - human
C:Species: Homo sapiens (man)
C:Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 05-Nov-1999
C:Accession: 138000; S51653
R:Parsons, S.F.; Mallinson, G.; Holmes, C.H.; Houlihan, J.M.; Simpson, K.L.; Mauby, W
Proc. Natl. Acad. Sci. U.S.A. 92, 5496-5500, 1995
R:Title: The Lutheran blood group glycoprotein, another member of the immunoglobulin
A:Reference number: 138000; MUID:95296337
A:Accession: 138000
A:Molecule type: mRNA
A:Residues: 1-628 <RES>
A:Cross-references: EMBL:X83425; NID:9603559; PIDN:CAA58449.1; PID:9603560
A:Note: parts of this sequence, including the amino end of the mature form, were confir
C:Genetics:
A:Gene: GDB:LU
A:Cross-references: GDB:120155; OMTM:111200
A:Map position: 19q12-19q13
C:Keywords: glycoprotein
F.1-31/Domain: signal sequence #status predicted <SIG>
F.32-628/Product: Lutheran blood group glycoprotein #status experimental <MAT>

Query Match      19.3%; Score 647.5; DB 2; Length 628;
Best Local Similarity 29.5%; Pred. No. 6,9e-34;
Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;

QY      2  GLEPR-LVCAFLLAACCCCPRVAGVGEAEQAPPELVNEYVSGTALLKGLSOSQGNLSH- 59
DB      12  GARRLLLLVLLA-----HPDAQEVRLLSVPLVEYWRKGSVLLD---TPGTGHDY 62
QY      60  -YVWF-----SVKKEKRTLEPRVRGOGQSEPGEGEQRSLSDRGATLALQV 106
DB      63  MLEWFLITDSCGARPRPLASLDEMOSSELOVTHMTBGRSP-----YLDQSGR---LVLAEA 115
QY      107  TPQDERIFLC--QGKPRSQEYRIQLRYVKADEPNIOVNLPLGIPVNSKEPEVATCYGR 164
DB      116  QVQDEEDYQCVVACAGAGTAETARLNVFAKPEATEVSPNKGTSLVMEDSAQELATCNRSR 175
QY      165  NGEPRIQVQVILWYKNGRKL--EENKRVHIQSOTV--ESSGLTYLOSILKLAOLYKEDKDAQ 221
DB      176  NGMPARKITWYRNGQRLERLVPEKNBPGVITSTRVBSAGLSLTITLVLRLKKDDRSNF 235
QY      222  YCELNRLRPSGNIMK--ESREVTVPVYPTPEKY--ML--EVEPVMILKSGDRVEIRCLADG 276
DB      236  HCAAHSLSEPGRGRGLDSTFHLTLHYPRHNQFWNGSGSTPAWGVRGGDVIQQLCRDGD 295
QY      277  NPPPHSISKQNPSTREAEETTNDN--GVLVLEPARKHSGRYVCCGLDIDTMI SL--LS 333

```

Db 296 SPSEPTTLFR-----LQDEOEVLNVLNLEGNLTLEGVTRGSGTGTCYGRVEDYDAADADVQLS 351  
QY 334 EPOELLVNVSDVRSVPAAPERQEG-----SLLTTEAEASSDLEFQWIREETGOV 385  
Db 352 KTELELRATVLDPELELS-----EGKVLSLPENNAAVNCVHGILPTPALRMKTDST--P 402  
QY 386 LERGPVLQDLHLKREAGGGRVCAVSPISIPGLNRTQLVNVAIFGPPMAFERK-----V 440  
Db 403 LQDGPMLSLSTTFDSNGTYCEASLPTVPVLSRTQNTLLVQSGPELKTAEIEPKADGS 462  
QY 441 WKENMVLNLSCEASGHPRTTSMNVNGTASEDDQDPOR---VLSTLVNLTVPPELLETGV 497  
Db 463 W-REDEVTLLICSAHGHPDKLSWSQLG--GSPAEPFIPRGQWVSSLLTKVTSLSRDGI 520  
QY 498 ECTASNDLQKNTSLILELVNLTTLTPDSTNTTGLSTSTASPTRANSTETKRLPEPES 557  
Db 521 SCEASNPHGKRRHVFHFGAAS-----PQTSQ 546  
QY 558 RGVVAVAVIYCIILVAVLGAVALYFLYKKGLPCRSGKOEITLPPSRKSELVEYVKSQKL 617  
Db 547 AGVAVMAVAVSVGLLLLVAVFYCVRRKGG--PCCRQRRKGAAPP---GEPGLSHSGSEQ 602  
QY 618 PEEMGLIQQ--SSGDKRAPGDGCK 640  
Db 603 PEQTGLIMGASGARGSGSGFGDE 627

RESULT 4  
137202  
B-CAM protein - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000  
C:Accession: I37202; S47272  
R:Campbell, I.G.; Roulees, W.D.; Senger, G.; Trowsdale, J.; Garin-Chesa, P.; Rettig, W.D.  
Cancer Res. 54, 5761-5765, 1994  
A:title: Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers:  
A:Reference number: I37202; MUID:95042297  
A:Accession: I37202  
A:status: preliminary; translated from GB/EMBL/DBJ  
A:molecule type: mRNA  
A:Residues: 1-588 <RES>  
A:cross-references: EMBL:X80026; NID:9535178; PIDN:CAAS6327.1; PID:9535179  
C:genetics:  
A:gene: B-CAM

Query Match 18.0%; Score 606; DB 2; Length 588;  
Best Local Similarity 29.2%; Pred. No. 2.9e-31;  
Matches 187; Conservative 96; Mismatches 247; Indels 110; Gaps 25;  
QY 2 GLPR-LVCAFLLAACCCCPRVAGVPEAEQAPAPELVEVVGSTALLKGLSOGNLSH- 59  
Db 12 GAFRLLLAVLLAA-----HPDAQAEVRLSVPLVEVNRKGSVLLDC---TPGTGHHY 62  
QY 60 -VDMF-----SVHKEKRLIFRVROGQSGSEPEYEOURLSLQDRGATLALQV 106  
Db 63 MLEWFLTRDSGARPLASAEQSGELQVMTMDLGRSP---YQDLSGR---LVLAFA 115  
QY 107 TQODEEIFLC--QGRKRQOEYRIQLRYKAPKEEPRIQVNPICITVYNSKEPEVATCYGR 164  
Db 116 QVQDEEDVYCVVACAGTAETARLNFAPKEATEVSPNKCTLLEVMEDSAOEIATCNSR 175  
QY 165 NGVPIPOVIMYKNGRPLK--EKKNRVHIQSSQTV--ESSGLYTLQSLTKAQLVKEKDKNOF 221  
Db 176 NGNPAPKIMYVNGQRLVPEVMPNPGVMTSRITVREASGLSLSTLTLVLPCKKDDRDASF 235  
QY 222 YCELNRLPLSGNHNK--ESSEVTVPVFVPEYK--WL--EVEPVGMKKGDRVEIQLADG 276  
Db 236 HCAAHYSLPEGHGRGLDSEPTFLTLHYPTREHYQFWGSPSPFAGVWRBDGVQLLCRGDG 295  
QY 277 NPPPHSISKQNPSTREAEETTNQ--GVVLPEPARKHSGRYCCGGLDLDTMISLSLSE 334  
Db 296 SPSEPTTLFR-----LQDEOEVLNVLNLEGNLTLEGVTRGSGTGTCYGRVEDYDA----- 344

QY 335 POELLVNVSDVRSVPAAP-ERQEG-----SLLTTEAEASSDLEFQWIREETGOV 385  
Db 345 ADDVOELSKTLIDVRAVYLDPELESEKVLSLPLNNSAAVNCVHGILPTPALRMKTDST--P 402  
QY 386 LERGPVLQDLHLKREAGGGRVCAVSPISIPGLNRTQLVNVAIFGPPMAFERK-----V 440  
Db 403 LQDGPMLSLSTTFDSNGTYCEASLPTVPVLSRTQNTLLVQSGPELKTAEIEPKADGS 462  
QY 441 WKENMVLNLSCEASGHPRTTSMNVNGTASEDDQDPOR---VLSTLVNLTVPPELLETGV 497  
Db 463 W-REDEVTLLICSAHGHPDKLSWSQLG--GSPAEPFIPRGQWVSSLLTKVTSLSRDGI 520  
QY 498 ECTASNDLQKNTSLILELVNLTTLTPDSTNTTGLSTSTASPTRANSTETKRLPEPES 557  
Db 521 SCEASNPHGKRRHVFHFGAAS-----PQTSQ 546  
QY 558 RGVVAVAVIYCIILVAVLGAVALYFLYKKGLPCRSGKOEITLPPSRKSELVEYVKSQKL 617  
Db 547 AGVAVMAVAVSVGLLLLVAVFYCVRRKGG--PCCRQRRKGAAPP---GEPGLSHSGSEQ 602

RESULT 5  
JH0506  
adhesion molecule SC1 precursor - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 21-Jul-2000  
C:Accession: JH0506; PS0270  
R:Tanaka, H.; Matsui, T.; Agata, A.; Tomura, M.; Kubota, I.; McFarland, K.C.; Kohr, B.  
Neuron 7, 535-545, 1991  
A:title: Molecular cloning and expression of a novel adhesion molecule, SC1.  
A:Reference number: JH0506; MUID:92030150  
A:Accession: JH0506  
A:Residues: 1-588 <TAN>  
A:cross-references: GB:S63276; NID:9238000; PIDN:AMB20170.1; PID:9238001  
A:Experimental source: embryo  
A:Accession: PS0270  
A:molecule type: protein  
A:Residues: 34-48 <TANI>  
C:Comment: This protein is uniquely and transiently expressed on spinal cord motoneur  
C:Keywords: glycoprotein; transmembrane protein  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:34-588/Product: adhesion molecule SC1 #status predicted <ADH>  
F:500-523/Domain: transmembrane #status predicted <TRA>  
F:101,173,199,271,312,366,462,485,504/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 15.2%; Score 509.5; DB 2; Length 588;  
Best Local Similarity 26.4%; Pred. No. 4.3e-25;  
Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;  
QY 6 LVCAFLLAACCCCPRVAGVPEAEQAPAPELVEVVGSTALLKGLSOGNLSHVM-FS 64  
Db 18 LLC--LLLAALCMPPALGL-----YYNNAVYCGITIMPCRLVLPDQ-LAFGKKIYE 65  
QY 65 VHKERKTLI-IR--VRQGGQSGSEPEYEOURLSLQDRGATLALQVTPQDERIFLCQ--CK 119  
Db 66 MPNSPVFIARSTTKRNVQYDVPDYKRLSLSE--NYTSLIKNARISDEKRRVCMVTE 124  
QY 120 RPRQOEYRIQLRYKAPKEEPRI--QVNPPLGIPVNSKREPEVAVCVGNGCPITQVIMYKN 177  
Db 125 DVVSEEPVY-VKVRQPSQPEILHQAFLL-----ETELKMLGECVNVDSYPEGNVIMYKN 179  
QY 178 GRPLK--EKKNRVHIQSSQTV--ESSGLYTLQSLTKAQLVKEKDKDAQFCELYNRLPSGNHM 235  
Db 180 GRVLOPVEEVVIVLKRVEN--RSTGLFTMTSSIQYMPKTKDAKAKFTCIYTYHGSPGQRT 238  
QY 236 KESKEVYVPPVYPEKAWLEY--EPVGMKKGDRVEIRCLADGNPPPH---FSTISKQNPST 291  
Db 239 IQSEPVVEDVHYPEKVTIRVLSQSSPTIKGDNVTLKSGNGNPPPOEFLFYI----- 291  
QY 292 REAEETTNQNGVILPEPARKHSGRYCCGGLDLDTMISLSPEQLLVNVSDVRSVPA 351

Db 292 -PGETGIRSSDPTVMTDVRNRATNGEYKCSLIDKSM-----DATITVHYL-DLQLPFS 344  
Qy 352 AP-EROGSSLTITLCEASSQDLFEQWLEETGOVLEREPVLQHLDKREAGGRCVAS 410  
Db 345 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOQODAGNICETT 400  
Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470  
Db 401 LOEVEGLKRRKTKLIVEGRP-QIMTKTKNTNKKSKITIVCHEGFPKPAVQMTVTS 458  
Qy 471 S--EODODPQVRLSTLN--VLVTPLELETGVECTASNDLGKNTSILFLELVNLTLLTPDS 526  
Db 459 SLINKTEETKYVNGKFSKRIIAPEENVTLTICIAENEL-----E 497  
Qy 527 NTTGLSTSTAS-----PHTRANSTSTERKLEPPESRGVIVAVIYCIILVALVAVLY 580  
Db 498 RIVTSLNVAISAIPEYDEPEDR--NDONSEK---VNDQAKLIVGIVGLLVALVAVLY 552  
Qy 581 FLY-KKGLPCRRSGKOEITLPPSRKSE 607  
Db 553 WLTVKSKTASKAVHDKDLGNIEENKLE 580

RESULT 6  
A45254  
surface glycoprotein BEN precursor - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 21-Jul-2000  
C:Accession: A45254, S19202  
R:Pouquile, O.; Corbel, C.; Le Caer, J.P.; Rossier, J.; Le Douarin, N.M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 5261-5265, 1992  
A:title: BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in  
A:reference number: A45254; MUID:92302224  
A:Accession: A45254  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-588 <POU>  
A:cross-references: EMBL:X64301; NID:g63087; PIDN:CAM45579.1; PID:g63088  
C:keywords: glycoprotein

Query Match 15.0%; Score 505.5; DB 2; Length 588;  
Best Local Similarity 26.4%; Pred. No. 7.7e-25;  
Matches 166; Conservative 116; Mismatches 255; Indels 91; Gaps 30;  
Qy 6 LVCAFLAACCCCPVAGVGEAEQAPAEVVEVGSTALLKCGLSQSGNLSHVDW-FS 64  
Db 18 LLC--LLLAALCMPPALG-----YTVNAVYGDITMPCRLVEYDGLMEGKMKYE 65  
Qy 65 VHKERTLI-FR--VRGGGSGSEPEYEQRLSLDRGATLALTOVTPDERIFLCO--GK 119  
Db 66 MPNGSPVFIARSTKKNVQYDDVDYKDRLSLSE--NYTSLIKARIRHEKRFVCMVTE 124  
Qy 120 RPRSGEYRIQLVYKAPPEPNI--QVNPGLIPVNSKEPEEVATCGRNGYPIPOVIWYKN 177  
Db 125 DVVSEEPYV-VKVFQPSPELHQADFL---ETEKLMKLGECVAVRDSYRDEGNVWYKN 179  
Qy 178 GRPLK--EENRNVHIQSSQVSSGLYTLQSLKAQLYEKDKDAQFCELYNRLPSGNHM 235  
Db 180 GRVLPVEEVVNIINRKEN--RSTGLFTMTSSLOQYMPKEDANNAKFTCIYVYHGSGQKT 237  
Qy 236 KESREYTVVFPYPTKRWLEV--EPVGMLEKEDGRVETICLAGDNPPH--FSISKQNST 291  
Db 239 IQSEPVFVHYPTREKTIIVLSSQSTIKGDVNTLKCSGNPPEQFLFYI-----291  
Qy 292 REAEETTNQGVLEVPARKHSGRYECQGLDITMTISLSEPOLLVNVYSDVRSVA 351  
Db 292 -PGETGIRSSDPTVMTDVRNRATNGEYKCSLIDKSM-----DTTITVHYL-DLQLPFS 344  
Qy 352 AP-EROGSSLTITLCEASSQDLFEQWLEETGOVLEREPVLQHLDKREAGGRCVAS 410  
Db 345 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOQODAGNICETT 400

Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470  
Db 401 HKEVGLKRRKTKLIVEGRP-QIMTKTKNTNKKSKITIVCHEGFPKPAVQMTVTS 458  
Qy 471 S--EODODPQVRLSTLN--VLVTPLELETGVECTASNDLGKNTSILFLELVNLTLLTPDS 526  
Db 459 SLINKTEETKYVNGKFSKRIIAPEENVTLTICIAENEL-----E 497  
Qy 527 NTTGLSTSTAS-----PHTRANSTSTERKLEPPESRGVIVAVIYCIILVALVAVLY 580  
Db 498 RIVTSLNVAISAIPEYDEPEDR--NDONSEK---VNDQAKLIVGIVGLLVALVAVLY 552  
Qy 581 FLY-KKGLPCRRSGKOEITLPPSRKSE 607  
Db 553 WLTVKSKTASKAVHDKDLGNIEENKLE 580

RESULT 7  
JH0464  
DM-GRASP precursor - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 02-Sep-2000  
C:Accession: JH0464  
R:Burns, F.R.; von Kannen, S.; Guy, L.; Raper, J.A.; Kamholz, J.; Chang, S.  
Neuron 7, 209-220, 1991  
A:title: DM-GRASP, a novel immunoglobulin superfamily axonal surface protein that sup  
A:reference number: JH0464; MUID:91337449  
A:Accession: JH0464  
A:Molecule type: mRNA  
A:Residues: 1-587 <BUR>  
A:Experimental source: brain  
C:Comment: This is a cell surface glycoprotein.  
C:Comment: This protein is localizes to axons in the dorsal funiculus and ventral mid  
C:keywords: glycoprotein  
E:1-32/Domain: signal sequence #status predicted <SIG>  
E:33-587/Product: DM-GRASP #status predicted <DMG>  
F:67,198,270,311,365,461,484,503/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 14.9%; Score 502.5; DB 2; Length 587;  
Best Local Similarity 26.3%; Pred. No. 1.2e-24;  
Matches 165; Conservative 118; Mismatches 254; Indels 91; Gaps 30;  
Qy 6 LVCAFLAACCCCPVAGVGEAEQAPAEVVEVGSTALLKCGLSQSGNLSHVDW-FS 64  
Db 17 LLC--LLLAALCMPPALG-----YTVNAVYGDITMPCRLVEYDGLMEGKMKYE 64  
Qy 65 VHKERTLI-FR--VRGGGSGSEPEYEQRLSLDRGATLALTOVTPDERIFLCO--GK 119  
Db 65 MPNGSPVFIARSTKKNVQYDDVDYKDRLSLSE--NYTSLIKARIRHEKRFVCMVTE 123  
Qy 120 RPRSGEYRIQLVYKAPPEPNI--QVNPGLIPVNSKEPEEVATCGRNGYPIPOVIWYKN 177  
Db 124 DVVSEEPYV-VKVFQPSPELHQADFL---ETEKLMKLGECVAVRDSYRDEGNVWYKN 178  
Qy 178 GRPLK--EENRNVHIQSSQVSSGLYTLQSLKAQLYEKDKDAQFCELYNRLPSGNHM 235  
Db 179 GRVLPVEEVVNIINRKEN--RSTGLFTMTSSLOQYMPKEDANNAKFTCIYVYHGSGQKT 237  
Qy 236 KESREYTVVFPYPTKRWLEV--EPVGMLEKEDGRVETICLAGDNPPH--FSISKQNST 291  
Db 238 IQSEPVFVHYPTREKTIIVLSSQSTIKGDVNTLKCSGNPPEQFLFYI-----290  
Qy 292 REAEETTNQGVLEVPARKHSGRYECQGLDITMTISLSEPOLLVNVYSDVRSVA 351  
Db 291 -PGETGIRSSDPTVMTDVRNRATNGEYKCSLIDKSM-----DATITVHYL-DLQLPFS 343  
Qy 352 AP-EROGSSLTITLCEASSQDLFEQWLEETGOVLEREPVLQHLDKREAGGRCVAS 410  
Db 344 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOQODAGNICETT 399  
Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470

```

OY 41 GSTALLKCGJSGOGNLSHDMFESVHKERTLLFRVNGOGGOS----EPREYDORLSIO 96
Db 36 GDTLIICRDLDPQ-NLMEFGKMKKEKDGSPVFLAFSSTKKSVOYDDVPEYDORLNLSE 94
OY 97 RGATLALVQYTPDERIFLQO-GKRPSOEYRIQLRVYKAPBEPNIQVNLGAIYVSKBP 155
Db 95 -NYLLSISNMAISDEKREKFCVLYEDVWFAPRTIVKFKQPSKEIYSKALFL--ETEQ 151
OY 156 EYAVTCGVRNGYPIPOYIWKNGSPKLEENNRHIOSSQVES-SGLYTLQSLTKQVL 214
Db 152 KKLDDCISEBDSYPRGNTITWRNGVULPRLGAVLIFPKKMDPVYTLQYTLSTLEKTKR 211
OY 215 EKDQAEYCELNRLPSGNMKESREYTVFEPYTERKWLVEV-EPYGMLEKSGREVRICL 273
Db 212 ADIOMPFCSTVYVGPFGQKTIHSEQAFDIYVPTBQVTLQVLPKNALKEGNIITLKC 271
OY 274 ADGNRPFH---FSLSKNPSTREAEETITDNGVULVLPARKHSGRYEGGDLDTMTIS 330
Db 272 GNGPBPPEBELFYLPQGPBGIRSSNYTLLMD-----VRNATDQYKSLLDKXSMJA 323
OY 331 LLSPEOELLVNYSDVNVSPAP-EROEGSSLLTLCBESSQDLQWLKEETGOVLERG 389
Db 324 STA---ITVHYL-DLSLNSGEVTRIGDALPVYSCISASRNATVYWMKMDNR--LRSS 376
OY 390 PVLQHLHKKREAGGQYCVASVPSIRGLANTOLVYNVAIFCPPMMAKREKRVVYKEMVILN 449
Db 377 P--SESSLHYQDAGNYCETALQEBELKKRESLTLIVEKP--QIKMTKTDPSGLSKT 432
OY 450 LSCFASHPRPITISMVNWGTAS--EODQDPQ---RVLSITLVNLYPPELLTETVECTASN 503
Db 433 IICHVEGFPKPAQIOWTITGSGSVYINQEEPSYINGRYS--KIISPEEYVTLTQTAEN 489
OY 504 DLGNNTSILPLDELVNLTTLLPDSNTTGLSTASPHTRANSTSTERKLPDEPSRGVVI 563
Db 490 QLETVNLSVNSAISI---PE-----HDEADEISDNR-EKVNDQAKTLLV 530

```

```
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
```

C:Accession: T20992; T24733  
R:Stulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20992  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <WIL>  
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:FL15G9.4a  
A:Experimental source: clone F15G9  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24733  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <WIL>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:FL15G9.4a  
C:Genetics:  
A:Gene: CESP:FL15G9.4a  
A:Map position: X  
A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1  
; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5175;  
Best Local Similarity 22.1%; Pred. No. 3.6e-09;  
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAQOPAPLVEVEVGSTAL-----LKCGLSQSGNLSHVMFVSYKKEKRTLIERVROG- 79  
DB 1357 DVQEPPIILPSTQTNNAVVGDRVELKCYEASPP--ASVTWF-----RRGI 1401  
QY 80 -QGSGEYEQRLSLQDRGATLALQVTPQDERIFLCQGRPSQ--EYRQLRYKAPE 137  
DB 1402 AIGDTKG-----YVESDGTLTQASVSDATITTCASNPAGKAEANLQVTVIASPD 1455  
QY 138 --EPNIQVNLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRPL-----KEKN 186  
DB 1456 IKDDPVYQD---SIKESHPFSLCYPFNS--PLQISMTLNDKRLIDDKTSWKTSDDKR 1510  
QY 187 RVHIQSQGVESGLYTLQSLKAOLYKEDKDAQFYCELNRLPSGNHMKESR-EVTVPV 245  
DB 1511 KLHVFRAK-ITDSGVYKCA-----RNAAGGSKSFQVEVTVPL 1548  
QY 246 FYFTEKYLEVEPVGMLEKGDVREIRCLADGNPPHFSI---SKQNPSTREABEETND 301  
DB 1549 NLDESKYKKRV---FAKEEVEYTLGCPVSGFPVQIWMVVDGTVVEPGKKYKATLSND 1604  
QY 302 NGVLVLEPARKHSGRYECG-----LDLDTMISLSEPOELLVNVYSDVRVSPAPER 355  
DB 1605 GLTLHFDSYSVQKGNHCHVAQSKGNLIDIDVELSVLAVP---YGEDDNLEFV----- 1655  
QY 356 QEGSSLTITC--EAESSDLEFQWL-----REETGOVLEGRPVQLQHLKREAGGGR 406  
DB 1656 -LGKDISLSCDLQTESDDKTFVWSINGESDRPDNVQIPSDGHRLYTTDAKPENNCKYM 1714  
QY 407 CVASVPSIFGLNRTQLVNV---AIFGPPWMAFKERKVVWKNVNLNLSGASGHPRTIS 463  
DB 1715 CRYL-NSAGKAERTLTLDVLEPPVFER--VFENQKLIGNNPII-LQCCVTONPKPYAI 1770  
QY 464 MNVNGTSEDODDQVRLSTLNVLPPELLETG---VECTASNDLGKNTSILFLELVN 519  
DB 1771 WKIDGNVDKSMFLFDESLIRI-----EKL-TKSAQISCTAENKAGTASRDFFIGNIA 1825  
QY 520 TTLLTPDSNTTT 530  
DB 1826 PTFKNEGDETF 1836

RESULT 11

T43290  
hemiceftin precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T43290; T20993; T24734  
R:Vogel, B.E.; Hedgcock, E.M.  
submitted to the EMBL Data Library, June 1998  
A:Description: Hemiceftin is required for hemidesmosome mediated cell adhesion and ge  
A:Reference number: Z22396  
A:Accession: T43290  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5198 <VOG>  
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1  
R:Stulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20993  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <WIL>  
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:FL15G9.4b  
A:Experimental source: clone F15G9  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24734  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <WIL>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:FL15G9.4b  
C:Genetics:  
A:Gene: him-4; FL15G9.4b  
A:Map position: X  
A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184  
; 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 303  
; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5198;  
Best Local Similarity 22.1%; Pred. No. 3.6e-09;  
Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAQOPAPLVEVEVGSTAL-----LKCGLSQSGNLSHVMFVSYKKEKRTLIERVROG- 79  
DB 1357 DVQEPPIILPSTQTNNAVVGDRVELKCYEASPP--ASVTWF-----RRGI 1401  
QY 80 -QGSGEYEQRLSLQDRGATLALQVTPQDERIFLCQGRPSQ--EYRQLRYKAPE 137  
DB 1402 AIGDTKG-----YVESDGTLTQASVSDATITTCASNPAGKAEANLQVTVIASPD 1455  
QY 138 --EPNIQVNLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRPL-----KEKN 186  
DB 1456 IKDDPVYQD---SIKESHPFSLCYPFNS--PLQISMTLNDKRLIDDKTSWKTSDDKR 1510  
QY 187 RVHIQSQGVESGLYTLQSLKAOLYKEDKDAQFYCELNRLPSGNHMKESR-EVTVPV 245  
DB 1511 KLHVFRAK-ITDSGVYKCA-----RNAAGGSKSFQVEVTVPL 1548  
QY 246 FYFTEKYLEVEPVGMLEKGDVREIRCLADGNPPHFSI---SKQNPSTREABEETND 301  
DB 1549 NLDESKYKKRV---FAKEEVEYTLGCPVSGFPVQIWMVVDGTVVEPGKKYKATLSND 1604  
QY 302 NGVLVLEPARKHSGRYECG-----LDLDTMISLSEPOELLVNVYSDVRVSPAPER 355  
DB 1605 GLTLHFDSYSVQKGNHCHVAQSKGNLIDIDVELSVLAVP---YGEDDNLEFV----- 1655  
QY 356 QEGSSLTITC--EAESSDLEFQWL-----REETGOVLEGRPVQLQHLKREAGGGR 406  
DB 1656 -LGKDISLSCDLQTESDDKTFVWSINGESDRPDNVQIPSDGHRLYTTDAKPENNCKYM 1714  
QY 407 CVASVPSIFGLNRTQLVNV---AIFGPPWMAFKERKVVWKNVNLNLSGASGHPRTIS 463

Db 1715 CRVT-NSAKAERTLTLDVLEPEVPEP--VEFANOKLIGNNPII-LQCVQGNPKPTVI 1770  
 Oy 464 WNNVNGASEDODPOKRVLSLTVLVPPELTG----VCTASNDLGKNTSILFELVNL 519  
 Db 1771 WKIDGNDVDKSWLPEDSLSLRLT-----EKL-TGKSAQISCTAENKAGTSRDFIIONIA 1825  
 Oy 520 TTLTPDSNTTT 530  
 Db 1826 PTFKNEGDOET 1836

## RESULT 12

A49448  
 Irregular chiasm C-roughest precursor - fruit fly (Drosophila melanogaster)  
 N:Alternate names: irrec-roughest protein  
 C:Species: Drosophila melanogaster  
 C:date: 11-Nov-1994 #sequence\_revision 11-Nov-1994 #text\_change 07-May-1999  
 C:Accession: A49448; S34129  
 R:Ramos, R.G., P.: Iqbal, G.L.; Lichte, B.; Baumann, U.; Maier, D.; Schneider, T.; Brandst  
 Genes Dev. 7, 2533-2547, 1993  
 A:title: The irregular chiasm C-roughest locus of Drosophila, which affects axonal proje  
 A:reference number: A49448; MUID:94102535  
 A:Accession: A49448  
 A:Status: preliminary  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1-764 <RAM>  
 A:Cross-references: GB:LI1040; NID:9304790; PID:9304791; EMBL:Z21641; NID:9312985; PID:9  
 A:gene: flyBase:rst  
 A:Cross-references: FlyBase:FBgn0003285  
 C:Keywords: transmembrane protein

Query Match 8.0%; Score 270; DB 2; Length 764;  
 Best Local Similarity 20.7%; Pred. No. 1.3e-09;  
 Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;

40 VGSTALLKCGLSQSGNLSHV-DWFSVHKERTLIFRVNQGQSGSPGEGEORLSIODKG 98  
 Db 41 VCARVTLPCRVINKOGTLOMTKDDFGLTGRDISGFERYAMWGSDDEGY----- 90  
 Oy 99 ATLALQVTPODERIFLCQCK-----RPRSOEYRIOLRYKAPPEPNIOVNLGIPIVNSK 153  
 Db 91 -SLDIYPMALDDARQCGVSPGEGOPAIRSTFAGLTIVLVPKPKITQGVITYATADR 149  
 Oy 154 EPEEVAITCVGRNGYPIPOYIWKV-GR-----PLKEKKNVHIISSQTVSSGL 201  
 Db 150 KVE--IECVSVGGKPAAEITWIDGLGNVLTNDIEYVIPLPDQR----- 192  
 Oy 202 YTLQSLTKAQVKEKDAQFYCELNRPLPSGNHMKESREVTYPTFKVLEVEPYGM 261  
 Db 193 FTKASVLRLLPKKEHNTNFSCQAQ--NTADRTYKSAIRKRYKAP--VKNVNMGK 246  
 Oy 262 LKEG-----DVEIRCLADGNPP--HFSISKNPSTREAE 295  
 Db 247 LRGAGAGSVGAGGSGVHMSSTGRIVESHQVLECRADANPDSVRYKRWINDPEITIGGK 306  
 Oy 296 EETTNDNGVLVLEPARKHSGRECOGLDITMISLSEPOELVNVSDVRSAPAPER 355  
 Db 307 TE-----WVINRWTKFHDIAIVKC--EVQNSVGKSEDSFLDISYASFRQROSMDA 357  
 Oy 356 QEGSSSLTLCESASODLEFQWLREFTGOVLERGPVLQHLDKRKRAAGGYRCVAVSPSP 415  
 Db 358 DVGSVSLTCEVDSPQPIVMIQHPDSRVVGTSTNLT--SVSNFAGRYCKAVNPGYA 416  
 Oy 416 GLNRQLVNVAFIPPEMAFKERKVVKENVNLNLSCEASGHPRT--ISWNVG--TASE 472  
 Db 417 EISADAVYVLK--GSPALG--SQRTOYGLVGTARIECFRASSVPRARHVSMTNGGEISSE 473  
 Oy 473 QDOD-----PQVLSLTVLVPPELTGECTASNDLGKNTSILFELVNLTTLP 524  
 Db 474 SGHDYSILVDAVPGVSKTLIRDSQAYHYGKYNCTVVDYGNDAVLEIQLAKKRVSL-- 531

Oy 525 DSNITTGSLTASPHRTANSTIERKLEPEPSGVIVANIVICILVAVGLVFLYK 584  
 Db 532 -----LMTIVG-----ISVVAQLVLTLT--VVVY---- 555  
 Oy 585 KGKLPGRSGROEITLPPSRKSELVEYKSDKLEPMGLQSSGDRAPDQGEYIDL 644  
 Db 556 ---IKCKRRTK---LPP-----ADVISEHQITKNGVSCSLERQDRISN:SDL 597  
 Oy 645 R 645  
 Db 598 K 598

## RESULT 13

JC2457  
 vascular cell adhesion protein - pig  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:date: 15-Feb-1995 #sequence\_revision 05-Apr-1995 #text\_change 08-Oct-1999  
 C:Accession: JC2457  
 R:Tseng, Y.T.M.; Haskard, D.O.; Robinson, M.K.  
 Biochem. Biophys. Res. Commun. 201, 805-812, 1994  
 A:title: Cloning and expression kinetics of porcine vascular cell adhesion molecule.  
 A:reference number: JC2457; MUID:94271236  
 A:Accession: JC2457  
 A:Molecule type: mRNA  
 A:Residues: 1-538 <TSA>  
 A:Cross-references: EMBL:U08351; NID:9474382; PIDN:AAA21542.1; PID:9474383  
 C:Keywords: glycoprotein; transmembrane protein  
 F:497-517/Domain: transmembrane #status predicted <TM>  
 F:75,157,271,330,360/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 7.9%; Score 264.5; DB 2; Length 538;  
 Best Local Similarity 21.6%; Pred. No. 1.8e-09;  
 Matches 126; Conservative 103; Mismatches 234; Indels 119; Gaps 27;

Oy 31 PAPELVEYEVGSTALLKCGLSQSGNLSHVDFSVHKERTLIFRVNQGQSGSPGEGYEQ 90  
 Db 31 PEDKMI-AQIDSDASLTCSADPCSSLS-FSW-----RTQIDSPUNG----- 70  
 Oy 91 RLSDODRGATLALQVTPODERIFLC-----QGRPRSOEYRIOLRYKAPPEPNIOV 143  
 Db 71 KVTNGTSTSLVMPNVSEFNHSLTCTVSCGNLKER-----GIQVETISPRKDPETHW 124  
 Oy 144 NPL--GIPVNSKEPEEYATCVGRNGYPIP--QVIWYKNGRPKLEEK--NRVHIISSQTV 196  
 Db 125 SSLPEVCGKPVIVR-----CLVPDQVYVPEKLEIELLKDNHMSWQSQFLEIDIKSETEK 177  
 Oy 197 ESSGLYTLQSLTKAQVKEKDAQFYCELNRPLPSGNHMKESREVTYPTFKVLEVEY 256  
 Db 178 SLEFFT-----PTEEDIGKAIVCOATLIDGQSPSVVTPPEKKMOVYISPPDPV-ISV 228  
 Oy 257 EPGVGLKEGDEVEIRCLADGNPPHPSISKONPSTREAEETTNDNGVLVLEPARKHSG 316  
 Db 229 NPSTSLQDGDMMATCTSEGLPAPQISMSK--LDNGDQDLISGNATLTLIMRMDSG 285  
 Oy 317 RYECQGLDITMISLSEPOELVNV--SDVRSAPAPERQEGSSSLTLCESASODL 373  
 Db 286 IYVEGVN--PVGITNKREVELYQVAPRDTTISVNSS--TLEGSSVNMTCSSDGPAP 341  
 Oy 374 EFQW--LREFTGOVLERGPVLQHLDKRKRAAGGYRCVAVSPSPICINRQLVNVAFIP 430  
 Db 342 KILMSKRLRDGNLPLESENTTLTITRKMDSGIIYVEGI--NOAGINRKE-VELIIOQA 398  
 Oy 431 P-----WMAFKERKVVKENVNLNLSCEASGHPRTITSMNVNNGTASBODDPOKRVLSLTV 486  
 Db 399 PKDQLTAFPSBSKBEQDYIISCTC--GNVPTLT-----ILKKAETGDIYVLSTDS 450  
 Oy 487 LVT--PELLETV-ECTASNDLGKNTSILFELVNLTTLPDSNTTTGSLTASPHRT 542  
 Db 451 AYTIHRAIADAGVYECESKMEIG-----LQLRSITLDVAGRSN----- 490

OY 543 ANSTERKLEPEPESRGVIAVAVICLVAVLAVLAVLYLK 584  
DB 491 KDFSSSE-----LVLYCASSLIIPALGVITTFARK 521

## RESULT 14

A39712  
kinase-like protein klg precursor - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 08-Nov-1991 #sequence\_revision 08-Nov-1991 #text\_change 24-Sep-1999  
C/Accession: A39712  
R:Chou, Y.H.; Hayman, M.J.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4897-4901, 1991  
A:Title: Characterization of a member of the immunoglobulin gene superfamily that possi  
A:Reference number: A39712; MUID:91271300  
A:Accession: A39712  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1051 <CHO>  
A:Cross-references: GB:M63437; NID:q212235; PIDN:AAA48933.1; PID:q212236  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo  
C:Keywords: ATP  
F:775-1046/Domain: protein kinase homology <KIN>  
F:783-791/Region: protein kinase ATP-binding motif

Query Match 7.7% Score 259.5; DB 2: Length 1051;  
Best Local Similarity 20.1%; Pred. No. 9.4e-09;  
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

OY 29 EOPABELVEVSGTALLKCGLSQSGNLSHVD-----WF-----SVHK 67  
DB 119 KQPA-SAAEIQPSSTVVLRC-----HIDGHPRTWQMFPGDAPLPDGRGTYSVS 167  
OY 68 EKRLTRVNGOGQSGPEY-----EQLSLQDKRATIALIQ-----V 106  
DB 168 KERLTLR---GAGPDNGIYYCSARPRAVGSCSDNFTLNIIDSEFQAVVVPEDLI 224  
OY 107 TPODERIFLCO--GKRPSOEY-----RIQLR---VYK--- 134  
DB 225 TKNEAMFDQFAVPPPTQEMLEDESPITNRSTYFANGSLIITQVKRSGVYKCI 284  
OY 135 -----APEPNIOVNLGITVNSKEBEVATCYGRNGYPIPOV 172  
DB 285 HGQKALVYLKATRLAEIEMAFSPKVLTAOGHRY-----SCACPGVPTPOV 335  
OY 173 IWKY-----GRPKKRNRYIIOSSQTVESGLYTLQSLIKALQVKEDKAOFCYL 225  
DB 336 WMEKNOERVPAGRVYQEAQLV--TSTIEADAGITTCNAKA----- 378  
OY 226 NYRLPSGNHMKESREVTVPVYPTKVMLEVEPYGMLKEGDRVEIRCLADGNPPHFS 285  
DB 379 -----GEKKQELSTIV---ATVPKVMEMPKDQLESKPGYIHLSTKASIKPTVTW 427  
OY 286 KQNSTEAEETNDNGVLLVLEPAKKEHSGRYECQGLDITMSLSSEQELLVNV--- 342  
DB 428 RNVGISIEDSRFEISENGTLRINNVVEYDGTMYKC-----VSSTAGSLIEGARV 477  
OY 343 --VSDVRSPAPAPRQ---EGSSLTLCFAESSODLEFQMLREB---TCQVLERGVLQ 393  
DB 478 HVLKELKFTPPPOQLQCKMEKKEVYCSATGRKPTIOTKTKDSSLPESHVSRAGILS 537  
OY 394 LHLDKREAGGCGYCVASVPSIPGLNRTQLVNVALFGPPMAFK--ERRVVKENMYNLNS 451  
DB 538 FHAKSRSDSCNYCTIASNSPGCEIRATVQLVAVY---VTFKLEPPTVYCGHTAMFQ 593  
OY 452 CEASGHPPTISWVNGTASQDODPORVLTSLNVLTPLELETVCTASNDLGKFTSI 511  
DB 594 COAGGDVPHIQMK-----GKDKIILDPKSLLPRIQIMPGSLV---IYDVTTEDSGKYTCI 646  
OY 512 -----LPELVNLTLPDSDMTTGLSTSTASPTPRANSTERRKLEPEESRGV 560  
DB 647 AGNSCNITKHREAFLLVVDKPAEEDG-----PSSHTPYKMIOT-----TGL 688

OY 561 VIVAVICLVAVLAVLAVLYLKKGKLPCKRSKQETLPPSRKSEL----- 608  
DB 689 SYGAAYVITIT--VLG--LMFYCKR-----RRKAKRLKHPGEPEHECNGTLQNG 739

OY 609 --VVEKSDKLEPEMGLLOGSS 628  
DB 740 QTTAEIQEVALTNIGSSGAS 761

## RESULT 15

LUXNL  
neural cell adhesion molecule long domain form precursor - African clawed frog  
N:Alternate names: NCAM-180  
M:Contains: neural cell adhesion molecule, short domain form (NCAM-140)  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 22-Jun-1999  
C/Accession: S09600  
R:Krieg, P.A.; Sakaguchi, D.S.; Kintner, C.R.  
Molecular Acids Res. 17, 10321-10335, 1989  
A:Title: Primary structure and developmental expression of a large cytoplasmic domain  
A:Reference number: S09600; MUID:90098871  
A:Accession: S09600  
A:Molecule type: mRNA  
A:Residues: 1-1088 <KRIT>  
A:Cross-references: EMBL:M25696; NID:q214609; PIDN:AAA49909.1; PID:q214610  
A:Note: the authors translated the codon AAA for residue 970 as Leu  
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM  
C:Comment: Several forms of NCAM are produced by alternative splicing.  
C:Genetics:

A:Gene: NCAM  
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; 1  
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-1088/Product: neural cell adhesion molecule, long domain form #status predicted  
F:20-803/1050-1088/Product: neural cell adhesion molecule, short domain form #status  
F:20-705/Domain: extracellular #status predicted <EXT>  
F:34-95/Domain: immunoglobulin homology <IMM2>  
F:129-188/Domain: immunoglobulin homology <IMM2>  
F:149-153/Region: heparin binding #status predicted  
F:158-162/Region: heparin binding #status predicted  
F:225-284/Domain: immunoglobulin homology <IMM3>  
F:317-381/Domain: immunoglobulin homology <IMM4>  
F:413-475/Domain: immunoglobulin homology <IMM5>  
F:512-589/Domain: fibronectin type III repeat homology <FN3A>  
F:618-679/Domain: fibronectin type III repeat homology <FN3B>  
F:706-723/Domain: transmembrane #status predicted <TM>  
F:724-1088/Domain: intracellular #status predicted <INT>  
F:41-93,136-186,237-282,323-379,420-473/disulfide bonds: #status predicted  
F:219,310,341,417,443,472/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 7.6% Score 256; DB 1: Length 1088;  
Best Local Similarity 18.1%; Pred. No. 1.7e-08;  
Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;

OY 33 PELVEVSGTALLKCGLSQSGNLSHVDSEVHKERTLFRVNGOGQSGPEYERL 92  
DB 26 PDGGEISLESKFILC---QVSGEATDISWSPTKEIKVT-----QOOI 66  
OY 93 SL---QDRGATLALQVTPRODERIFLC-----QGR-----RPRSOEY 126  
DB 67 SVYRSDVYSTLTITYNASSQDAGITKCVASNEAGESEGTYNLKYOKLTRNAPTPDEF 126  
OY 127 R-----TOLRYKAPPE----- 138  
DB 127 KEEEDAVIICDVSSSTPITTRHKKGVIEKKDREFVLANNVYQINGIKTTDGTGTRC 186  
OY 139 -----PNIQVNDLGLIPVNSKEPEVATCGRNGPIPOVITYK 176  
DB 187 EGRILARGEINVKDIQVIVNVPPTIOARQLRVNATANNAEVSVLSCDADGFPDPEISWLK 246  
OY 177 NGRPLKEKNRHHIQSSQVSESSGLYTLQSLIKALQVKEDKAOQVYCELANTRLPNGNMK 236



```

Db 247 KGEPLDGEKISFNED-----QSEMTIHHEKDEAEYSICIAN-----NQAG 289
QY 237 ESREYTVVFPTEKVMLEVEPVMLEKGDREVEIRCLADNPPPHFISIKONSTREAE 296
Db 290 EAEATILIKYAKKITYVENKTAV--ELDEITLTCESGDDIP--SITWRTAVRNTISSE 345
QY 297 ETTNDNGVLVEPARKE-----HSGRYEC-----OGLDLDTMISLSEPOELLVN 341
Db 346 ATTLDGHIYVKEHIRMALITLKDQYTDAGEYFCIASNPIGVDMQAMV-----FEVQ 397
QY 342 YSDVRRVSPAAPEROGSSSLTTCESASODLEFOWLREFTGOVLE-----RGPV 391
Db 398 YAPKIR-GPVVVVYTWEGNPNVITCEVFAHPRAAVTFERD--GOLLPSSNFSNITKISGP 454
QY 392 ---LQLDLREAGGGRV-----ASVPSIGLNTQOLVN--VAIFG 429
Db 455 SSSLEVPDSENDENFGNCTAINTIGHFSEFLLVQADTPSPAIRKVEPYSSTVMIVFD 514
QY 430 PP-----WMAFERKRVWVK-----ENM--VLNLSCEAS-----455
Db 515 EPDSTGVPIILKYKAEMRVIGHEKMHKYYDAKEVNAESIITYMGLKETSYWKLAMN 574
QY 456 ----GHPRTISMNVN-----GTASE-----ODQDPORV-----LST 483
Db 575 GKGLGDSTPSQEFITQVREPSAPKLVGHLSEDSNSIKVDILKODDGGSPIRHYLVNRA 634
QY 484 LNVLV--TPEL-----LETGYE-----CTASNDLGKNTSL--FLELVNLTTLT 523
Db 635 LNALEKPEKRVPSNSHHVWLKALEWVDYEVIVAENOQKSPALLSFRTAKPTATT 694
QY 524 PDSNTTGLSTASPTRANSTSTERKLPESRGVIVAVICILVLAVALGYFLY 583
Db 695 ATASAGTGLGTG-----AIVGILIVIEVLLVVDVTCFFLN 731
QY 584 KKGKLC-----RRSGKOETLPPSRKSELVEVKSCLKPEMGLQG 626
Db 732 KCGILMCIAVFCGKAGPGAKGKDIEGKAAPFSKDESK--EPIVEVTE-----778
QY 627 SSGDKRAPGDGEXYID 643
Db 779 ---EERTPNHDSNQIE 792

```

Search completed: March 23, 2001, 12:31:19  
 Job time: 42 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:42 ; Search time 13.25 Seconds

(without alignments)  
1574.489 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGJPLVCAFLALACCCPR.....SSGDKRAPDQGEKTYDLRH 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3327	98.9	646	1	MU18_HUMAN
2	647.5	19.3	628	1	LU_HUMAN
3	509.5	15.2	588	1	C166_CHICK
4	492.5	14.6	583	1	C166_HUMAN
5	488.5	13.5	583	1	C166_MOUSE
6	465.5	13.8	555	1	C166_CARAU
7	438	13.0	564	1	C166_BRARE
8	270	8.0	764	1	ICCG_DROME
9	259.5	7.7	1051	1	PTK7_CHICK
10	256	7.6	1088	1	NCAL_XENLA
11	246.5	7.3	725	1	NCAL_MOUSE
12	246.5	7.3	739	1	NCAL_RAT
13	244.5	7.3	858	1	NCAL_RAT
14	244.5	7.3	3707	1	PGBM_MOUSE
15	242.5	7.2	853	1	NCAL_BOVIN
16	237	7.0	1092	1	NCAL_XENLA
17	235	7.0	4333	1	PGBM_HUMAN
18	235	7.0	1091	1	NCAL_CHICK
19	234	7.0	761	1	NCAL_HUMAN
20	234	7.0	848	1	NCAL_HUMAN
21	233.5	6.9	1115	1	NCAL_MOUSE
22	232.5	6.9	1260	1	CAML_MOUSE
23	230	6.8	739	1	VCAI_HUMAN
24	230	6.8	837	1	NCM2_MOUSE
25	228	6.8	811	1	FS22_DROME
26	228	6.8	873	1	FS21_DROME
27	226.5	6.7	1259	1	CAML_RAT
28	226	6.7	738	1	PECL_HUMAN
29	225.5	6.7	1257	1	CAML_HUMAN
30	225	6.7	702	1	CCEM_HUMAN
31	224.5	6.7	1070	1	PTK7_HUMAN
32	218.5	6.5	404	1	RAGE_HUMAN
33	215	6.4	1284	1	NRCA_CHICK

RESULT ID	1	MU18_HUMAN	STANDARD:	PRT:	646 AA.
AC	P43121;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	CELL SURFACE GLYCOPROTEIN MUC18 PRECURSOR (MELANOMA-ASSOCIATED ANTIGEN MUC18) (MELANOMA-ASSOCIATED ANTIGEN A32) (S-ENDO 1 ENDOTHELIAL-ASSOCIATED ANTIGEN) (CD146 ANTIGEN) (MELANOMA ADHESION MOLECULE).				
DE	MCAM OR MUC18.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=MELANOMA;				
RX	MEDLINE=90099368; PubMed=2602381;				
RA	Lehmann J.M., Riethmuller G., Johnson J.P.;				
RT	"MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily."				
RT	Proc. Natl. Acad. Sci. U.S.A. 86:9891-9895(1989).				
RL	[2]				
RN	REVISIONS, SEQUENCE FROM N.A.				
RP	TISSUE=MELANOMA;				
RX	MEDLINE=93391384; PubMed=8378324;				
RA	Sers C., Kirsch K., Rothbacher U., Riethmuller G., Johnson J.P.;				
RT	"Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains."				
RT	Proc. Natl. Acad. Sci. U.S.A. 90:8514-8518(1993).				
RL	[3]				
RN	SEQUENCE OF 24-44; 98-112; 135-153; 240-260; 379-389 AND 460-478.				
RP	MEDLINE=94215196; PubMed=8162602;				
RX	Shih I.-M., Elieder D.E., Speicher D., Johnson J.P., Herlyn M.;				
RA	"Isolation and functional characterization of the A32 melanoma-associated antigen."				
RT	Cancer Res. 54:2514-2520(1994).				
RL	[4]				
RN	SEQUENCE OF 27-40; 98-112 AND 236-260.				
RP	MEDLINE=96136502; PubMed=8573133;				
RX	Barlin N., Frances V., Lesaulle G., Horschowski N., George F., Sampol J.;				
RA	"Identification of the S-Endo 1 endothelial-associated antigen."				
RT	Biochem. Biophys. Res. Commun. 218:210-216(1996).				
RL	[5]				
RN	FUNCTION.				
RP	MEDLINE=94122526; PubMed=8292890;				
RX	Johnson J.P., Rothbacher U., Sers C.;				
RA	"The progression associated antigen MUC18: a unique member of the immunoglobulin supergene family."				
RT	Melanoma Res. 3:337-340(1993).				
RL	FUNCTION: COULD BE AN ADHESION MOLECULE ACTIVE IN NEURAL CREST CELLS DURING EMBRYONIC DEVELOPMENT. ITS EXPRESSION MAY ALLOW MELANOMA CELLS TO INTERACT WITH CELLULAR ELEMENTS OF THE VASCULAR SYSTEM THEREBY ENHANCING HEMATOGENEOUS TUMOR SPREAD.				

P20273 homo sapien  
P35969 mus musculu  
P35329 mus musculu  
P29533 mus musculu  
P13688 homo sapien  
Q63495 rattus norv  
P35968 homo sapien  
Q28173 bos taurus  
Q60469 homo sapien  
P03696 gallus gall  
P70211 mus musculu  
P97798 mus musculu

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN  
 CC CELLS DURING THEIR EMBRYONIC MIGRATION. APPEARS TO BE LIMITED TO  
 CC VASCULAR SMOOTH MUSCLE IN NORMAL ADULT TISSUES. ASSOCIATED WITH  
 CC TUMOR PROGRESSION AND THE DEVELOPMENT OF METASTASIS IN HUMAN  
 CC MALIGNANT MELANOMA. EXPRESSED MOST STRONGLY ON METASTATIC LESIONS  
 CC AND ADVANCED PRIMARY TUMORS AND IS ONLY RARELY DETECTED IN BENIGN  
 CC MELANOCYTIC NEVI AND THIN PRIMARY MELANOMAS WITH A LOW PROBABILITY  
 CC OF METASTASIS.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC -1- DATABASE: NAME-PROW: NOTE-CD guide CD146 entry:  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd146.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib.ch).  
 CC -----  
 DR EMBL: M29277: AAA20824.1: -  
 DR EMBL: M28882: AAA20932.1: -  
 DR EMBL: X68264: CAA48332.1: JOINED.  
 DR EMBL: X68265: CAA48332.1: JOINED.  
 DR EMBL: X68266: CAA48332.1: JOINED.  
 DR EMBL: X68267: CAA48332.1: JOINED.  
 DR EMBL: X68268: CAA48332.1: JOINED.  
 DR EMBL: X68270: CAA48332.1: JOINED.  
 DR EMBL: X68271: CAA48332.1: JOINED.  
 DR MIM: 155735: -  
 DR INTERPRO: IPR003006: -  
 DR PFM: PFO0047: 19: 5.  
 DR Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 646  
 FT DOMAIN 24 559  
 FT TRANSMEM 560 583  
 FT DOMAIN 584 646  
 FT DOMAIN 39 122  
 FT DOMAIN 154 229  
 FT DOMAIN 265 327  
 FT DOMAIN 358 414  
 FT DOMAIN 445 506  
 FT DISULFID 48 116  
 FT DISULFID 161 223  
 FT DISULFID 272 320  
 FT DISULFID 365 407  
 FT DISULFID 452 499  
 FT CARBOHYD 56 56  
 FT CARBOHYD 418 418  
 FT CARBOHYD 449 449  
 FT CARBOHYD 467 467  
 FT CARBOHYD 508 508  
 FT CARBOHYD 518 518  
 FT CARBOHYD 527 527  
 FT CARBOHYD 544 544  
 SQ SEQUENCE 646 AA; 71793 MW; F064A5DAE0BAEC6 CRC64;

Query Match 98.9%; Score 3327; DB 1; Length 646;  
 Best Local Similarity 98.9%; Pred. No. 9.3e-213;  
 Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLRLVCAFLLAACCCCPVAGVPGAEAPAPLVEVENSTALLCGLSOSQGNLSHV 60  
 DB 1 MGLRLVCAFLLAACCCCPVAGVPGAEAPAPLVEVENSTALLCGLSOSQGNLSHV 60  
 QY 61 DMFSVHKRKLIRFRVROGOGSPEGEORLSLDDGATATLQVTPDPRIFILCOGR 120  
 DB 61 DMFSVHKRKLIRFRVROGOGSPEGEORLSLDDGATATLQVTPDPRIFILCOGR 120

QY 121 PRSGEYRIOLRVYKAPPEPNIOVNLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRP 180  
 DB 121 PRSGEYRIOLRVYKAPPEPNIOVNLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRP 180  
 QY 181 LKEERNVHIQSSQTVSSGLYTLQSLKALVKEKDAOFYELNRLPSGHMKESRE 240  
 DB 181 LKEERNVHIQSSQTVSSGLYTLQSLKALVKEKDAOFYELNRLPSGHMKESRE 240  
 QY 241 VTPVFPTEKVMLEVEPVGMLKEGDRVEIRCLADGNPPHFSISKONPSTRAEETTN 300  
 DB 241 VTPVFPTEKVMLEVEPVGMLKEGDRVEIRCLADGNPPHFSISKONPSTRAEETTN 300  
 QY 301 DNGVLVEPARKESGREGCGLDLDTMISLSEPOELLVNVYSDVSPAPAREQESS 360  
 DB 301 DNGVLVEPARKESGREGCGLDLDTMISLSEPOELLVNVYSDVSPAPAREQESS 360  
 QY 361 LTLTCEASSODLEFQWLREFTGOVLERGPLYLQHLDKREAGGGRVAVSPSTPLGNT 420  
 DB 361 LTLTCEASSODLEFQWLREFTGOVLERGPLYLQHLDKREAGGGRVAVSPSTPLGNT 420  
 QY 421 QLVNVAITPGPMWAFKEKRYVKNVNLNISCASGHPPTISNVNGTASEODDOPQV 480  
 DB 421 QLVNVAITPGPMWAFKEKRYVKNVNLNISCASGHPPTISNVNGTASEODDOPQV 480  
 QY 481 LSTLVNLTPELLEFGVECTASNDLGKNTSLFLELVNLTTPDSTMTTGLSTVASPH 540  
 DB 481 LSTLVNLTPELLEFGVECTASNDLGKNTSLFLELVNLTTPDSTMTTGLSTVASPH 540  
 QY 541 TRANSTERKLPPEESGVYIVAVICILVAVLGVAVLYLYKKGKLPGRSKOETTL 600  
 DB 541 TRANSTERKLPPEESGVYIVAVICILVAVLGVAVLYLYKKGKLPGRSKOETTL 600  
 QY 601 PPSRSELVEYKSDLPPEEMGLLOGSGDKRAPDQGEKTYIDLRH 646  
 DB 601 PPSRSELVEYKSDLPPEEMGLLOGSGDKRAPDQGEKTYIDLRH 646

RESULT 2  
 ID LU HUMAN STANDARD: PRT; 628 AA.  
 AC P50895:  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE LUTHERAN BLOOD GROUP GLYCOPROTEIN PRECURSOR (B-CAM CELL SURFACE  
 DE GLYCOPROTEIN) (AUBERGER B ANTIGEN) (F8/G253 ANTIGEN).  
 GN LU OR BCAM OR MSK19.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.  
 RC TISSUE-PLACENTA;  
 RX MEDLINE-9596337; PubMed-7777537;  
 RX Parsons S.F., Mallinson G., Holmes C.H., Houlihan J.M., Simpson K.L.,  
 RA Wadby W.J., Spurr N.K., Wayne D., Barclay A.N., Anstee D.J.:  
 RT "The Lutheran blood group glycoprotein, another member of the  
 RT immunoglobulin superfamily, is widely expressed in human tissues and  
 RT is developmentally regulated in human liver.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).  
 RN [2]  
 RP SEQUENCE OF 1-588 FROM N.A.  
 RX MEDLINE-95042297; PubMed-7954395;  
 RX Campbell I.G., Foulkes W.D., Senger G., Trowsdale J.,  
 RA Garbin-Chesa P., Retlig W.J.:  
 RT "Molecular cloning of the B-CAM cell surface glycoprotein of  
 RT epithelial cancers: a novel member of the immunoglobulin  
 RT superfamily.";  
 RL Cancer Res. 54:5761-5765(1994).  
 CC -1- FUNCTION: PROBABLE RECEPTOR. MAY MEDIATE INTRACELLULAR SIGNALING.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION (HIGHEST IN THE

CC PANCREAS AND VERY LOW IN BRAIN). CLOSELY ASSOCIATED WITH THE BASAL  
 CC LAYER OF CELLS IN EPITHELIA AND THE ENDOTHELIUM OF BLOOD VESSEL  
 CC WALLS.  
 CC -1- DEVELOPMENTAL STAGE: IS UNDER DEVELOPMENTAL CONTROL IN LIVER AND  
 CC MAY ALSO BE REGULATED DURING DIFFERENTIATION IN OTHER TISSUES.  
 CC UPREGULATED FOLLOWING MALIGNANT TRANSFORMATION IN SOME CELL TYPES.  
 CC -1- POLYMORPHISM: LD IS RESPONSIBLE FOR THE LUTHERAN BLOOD GROUP  
 CC SYSTEM.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X83425; CAA58449.1; -  
 DR EMBL; X80026; CAA56327.1; -  
 DR MIM; 111200; -  
 DR INTERPRO; IPR003006; -  
 DR PFM; PF00047; 1g; 5.  
 KW Receptor; Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;  
 KW Blood group antigen.  
 FT SIGNAL 1 31  
 FT CHAIN 32 628 LUTHERAN BLOOD GROUP GLYCOPROTEIN.  
 FT DOMAIN 32 547 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 548 568 POTENTIAL.  
 FT DOMAIN 569 628 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 46 132 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 165 244 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 284 344 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 377 431 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 466 529 IG-LIKE C2-TYPE DOMAIN.  
 FT DISULFID 53 125 PROBABLE.  
 FT DISULFID 172 237 PROBABLE.  
 FT DISULFID 291 337 PROBABLE.  
 FT DISULFID 384 424 PROBABLE.  
 FT DISULFID 473 522 PROBABLE.  
 FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 225 226 RL -> PC (IN REF. 2).  
 FT CONFLICT 355 356 EL -> DV (IN REF. 2).  
 SQ SEQUENCE 628 AA; 67374 MW; C89B0A4835492B1E CRC64;

Query Match 19.3%; Score 647.5; DB 1; Length 628;  
 Best Local Similarity 29.5%; Pred. No. 1.5e-35;  
 Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;

QY 2 GUR-LVCAALLAACCCCPVAVGPEAEPAPELVEVEGSTALLKIGISQSGMLSR- 59  
 DB 12 GABPLLLAVLLAA-----HPDAQAEVRLSVPLPEVMGKGVILDC--TPTGTHDH 62  
 QY 60 -VWME-----SVHKEKRTLLIFRYRGOGSGSEPEYQRSLDQKGTALTQV 106  
 DB 63 MLEMFLLDRSGARPLASAMQSELOVTHHTGRSP---YOLDSGR--LVLA 115  
 QY 107 TPQDERIFLC--QGRKRSQERYRIOLRVYKAPPEPNIQVPLGIPNRSKEPEVATCVR 164  
 DB 116 QVDEDERYVAVRAGAAGTATARNLVFAKPEATEVSPKGTLSWEDSACDIATCNSR 175  
 QY 165 NGPIPIQVIVYKGRPLK--EENRNVHIOSSQTV--BSSGIYTLQSLKQVLEKDAQF 221  
 DB 176 NGNPAPITWYNGORLEVEVEMNPEGYMTSRVREASGLISTFLYRLRLKRDQDASF 235  
 QY 222 YCELNRLPSGNHMK--ESREVTVPVPEYPEKY--WL--EVEPYGMKEDREIRCLADG 276

DB 236 HCAAHYSLEPGRGRGLDSPFHLTLHYPTHEHGFVWVSGSPSTRAGVREGDTYOLLCRDG 295  
 QY 277 NPPPHSISKQNSTEAEETTNDN--GVLYEPARKHSGRYECQGLDITMISL-LS 333  
 DB 296 SPSEYTLFR---LQDEQEVLTNMLEGNTLEGTWQSGTYGCRVEDYDAADVOLS 351  
 QY 334 EPELLVNVSDYRVAPAPEREQ-----SSLTLCGEASSDLEFQVLRRETCV 385  
 DB 352 KTEFLRAYATDPLELS-----EGKVLSPLNSSAVVNCVHGILPTPALRTKST--P 402  
 QY 386 LERGPVLOLHLDRKREAGGRCVAVSPISPIGNTQLVAVAFGPPMAFERK-----V 440  
 DB 403 LGGPMILSSITFDEPNSGTVCASLPYVPLSRQNTFLVQGSSELPKTAIEPKAGS 462  
 QY 441 WVENNVNLNCSASCHPPTPTSMWNVNGTASQDDQPR---VLSTLVNLYTPELLETV 497  
 DB 463 W-REGDEVLTISARHPKPKLSWSQLG--GSPAEPFPGQVWSSSLTLKVSALSROI 520  
 QY 498 ECTASNDLCKNTSILFLELVNLTTLTPDSNTTGLSTASPHITRANSTSRKLPPEES 557  
 DB 521 SCASNPNGKRRVHFEGAVS-----PQTSQ 546  
 QY 558 RGVVIVAVTICILVAVLAVGLVFLYKKGKLPQRSKQETLPPSRKSELVVEYKSDKL 617  
 DB 547 AGVAVMAVAVSVGLLLVAVFVYVARRKG--PCQRQREKAPPP---GEGLSHSGSEQ 602  
 QY 618 PREMGILQ--SSGDKRAPDQGEK 640  
 DB 603 PRQGTGLMGASGAGSGSGFGDE 627

RESULT 3  
 ID C166-CHICK STANDARD; PRT; 588 AA.  
 AC P42292;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CD166 ANTIGEN PRECURSOR (SCL GLYCOPROTEIN) (BEN GLYCOPROTEIN) (DM-  
 DE GRASP PROTEIN) (JC7 PROTEIN).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-53.  
 RC TISSUE=EMBRYO;  
 RA MEDLINE=92030150; PubMed=1931049;  
 RA Tanaka H., Matsui T., Agata A., Tomura M., Kubota I.,  
 RA McFarland K.C., Kohr B., Lee A., Phillips H.S., Shelton D.L.;  
 RT "Molecular cloning and expression of a novel adhesion molecule, SCL.",  
 RL Neuron 7:535-545(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=91337449; PubMed=1873027;  
 RA Burns F.R., von Kannen S., Guy L., Raper J.A., Kamholz J.,  
 RA Chang S.;  
 RT "DM-GRASP, a novel immunoglobulin superfamily axonal surface protein  
 RT that supports neurite extension.",  
 RL Neuron 7:209-220(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-52; 87-93; 504-512 AND 569-582.  
 RC TISSUE=BURSA OF FABRICIUS;  
 RA MEDLINE=92302224; PubMed=1608932;  
 RA Pourquie O., Corbel C., le Caer J.-P., Rossier J., le Douarin N.M.;  
 RT "BEN, a surface glycoprotein of the immunoglobulin superfamily, is  
 RT expressed in a variety of developing systems.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5261-5265(1992).  
 RN [4]  
 RP POSSIBLE FUNCTION.  
 RA MEDLINE=92211411; PubMed=1313497;  
 RA Pourquie O., Hallonet M.E.R., le Douarin N.M.;  
 RT "Association of BEN glycoprotein expression with climbing fiber

RT axonogenesis in the avian cerebellum.";  
 RL J. Neurosci. 12:1548-1557(1992).  
 CC -1- FUNCTION: HOMOPHILIC ADHESION MOLECULE ASSOCIATED WITH CLIMBING  
 CC FIBER AXONEMESIS. SUPPORTS NEURITE EXTENSION.  
 CC -1- CELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A RESTRICTED POPULATION OF AXONS.  
 CC WITHIN THE SPINAL CORD IT IS LOCALIZED TO AXONS IN THE DORSAL  
 CC FUNICULUS, MIDLINE FLOOR PLATE CELLS, AND MOTONEURONS. FOUND IN  
 CC EPITHELIA, CENTRAL NERVOUS SYSTEM, AND HEMOPOIETIC CELLS. EACH  
 CC DEVELOPING SYSTEM CARRIES DIFFERENTLY GLYCOSYLATED MOLECULAR FORMS  
 CC OF BEN.  
 CC -1- DEVELOPMENTAL STAGE: WIDELY EXPRESSED DURING EMBRYONIC  
 CC DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S63276; AAB20170.1; -  
 DR EMBL: M76678; AAA48602.1; -  
 DR EMBL: X64301; CAA45579.1; -  
 DR HSSP: Q13740; IKCC.  
 DR INTERPRO: IPR000495; -  
 DR INTERPRO: IPR003006; -  
 DR PIRAM: PF00047; 19; 5.  
 DR PROSITE: PS00280; IG\_MHC; FALSE\_NEG.  
 DR Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;  
 KM Signal.  
 FT 1 33  
 FT STGNAL 34 588  
 FT CHAIN 34 532  
 FT DOMAIN 533 553  
 FT TRANSMEM 554 588  
 FT DOMAIN 42 126  
 FT DOMAIN 156 233  
 FT DOMAIN 269 326  
 FT DOMAIN 352 404  
 FT DOMAIN 433 497  
 FT DISULFID 49 119  
 FT DISULFID 163 226  
 FT DISULFID 276 319  
 FT DISULFID 359 397  
 FT DISULFID 440 490  
 FT CARBOHYD 101 101  
 FT CARBOHYD 173 173  
 FT CARBOHYD 199 199  
 FT CARBOHYD 271 271  
 FT CARBOHYD 312 312  
 FT CARBOHYD 366 366  
 FT CARBOHYD 462 462  
 FT CARBOHYD 485 485  
 FT CARBOHYD 504 504  
 FT CONFLICT 2 10  
 FT CONFLICT 25 25  
 FT CONFLICT 112 113  
 FT CONFLICT 329 329  
 FT CONFLICT 401 402  
 FT CONFLICT 401 402  
 FT SEQUENCE 588 AA; 65726 MW; 2A28612D0164531E CRC64;

Query Match 15.28; Score 509.5; DB 1; Length 588;  
 Best Local Similarity 26.48; Pred. No. 1.8e-26;  
 Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;

QY 6 LVCAFLAACCCPRVAGVPEAGPAPELVVEVSTALCKGLSQSGNLSHVDW-FS 64  
 DB 18 LLC--LLLAALCMPALGL-----YTVNAVYGDITTPCRLEVPDGLMFGKKKYE 65

QY 65 VHKERTLI-FR--VRQGGQSEPEXEQRLSDRGATLALQVTPDERIFLCQ--GK 119  
 DB 66 MNPSPVLIARRSSKRVQDDPDYDKRLSIE-NYTLISIKARISDERKFCMLYTE 124  
 QY 120 RPRSGEYRIOLRVYKAPPEPNI-QVNPGLIPVNSKEPEEYAVCVGNGYPIPOVIYKN 177  
 DB 125 DVSEEPVY-VKVEKQPSPELHQADEL----ETEKLMIGECVVRDSYEGNVTYKN 179  
 QY 178 GRPLK--EKKRVHIIQSQYVESSGLTYLQSLIAQLVKEKXDAQFCELYRLPSGNHM 235  
 DB 180 GRVLQPVVEEVVYINLRKEN-RSTGLFTYTSLOYMPTKEDANKFCYLYHGPSQKT 238  
 QY 236 KSRREVTPVPEPTREKMLEV-EPYGMLESGDRVEIRCLADGNPPH---FSISKQNPST 291  
 DB 239 IQSEEVVVDVHYPTPKVITIRLSQSTIKEDNNTLKSSGNGNPQPELFYI----- 291  
 QY 292 REAEETNDNGVLEVPAREKHSRGYCGQLDPMISLSEPOELLVTVSVRSPA 351  
 DB 292 -PGETEGIRSDPTVYMTVRRNATGEYKCSLIDKSM-----DATYTVHYL-DLQLTPS 344  
 QY 352 AP-EROEGSSLLTNGEASSDLEFQWMLREFGVLEGPVLOLHDLKREAGGYRCVAS 410  
 DB 345 GEVTQIGALPVSCTTSSSRNATYFWIKDNR--MKTSP--SFSSIQYQAGNYICETT 400  
 QY 411 VPSIGLNRTOLVNVAIFGPPMAFERKRVYKEMVNLNLSCEASGHPRTISMNVNGTA 470  
 DB 401 LQEVGLAKRKRTLKLIVGKP--QIKMTKNTNMKSTVICHVEGFPAVQWTVVGS 458  
 QY 471 S--EODQDQRLSTLN--VLATPELLETGYECTASNLGKNTSLFELVNLTLTPDS 526  
 DB 459 SLINTEKEKYVNGKFSKIIIAPEENVY-LTCLAEENL-----E 497  
 QY 527 NNTTGLSTSTAS-----PHTRANSTTERKLPEPESRGVYIVAVICLVAVGAVLY 580  
 DB 498 KVTSLNSAISIPREYDEEDR-NDNSEK-----VNDQAKLIVGLVGLLVALVAGVY 552  
 QY 581 FLY-KKGLPCRRSGKQETLPPSRKE 607  
 DB 553 WLYVKKSTASKVHVDKLDGNIENKLE 580

RESULT 4  
 C166\_HUMAN STANDARD; PRT; 583 AA.  
 AC Q13740; O60892;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CD166 ANTIGEN PRECURSOR (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE)  
 DE (ALCAM).  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95279947; PubMed=7760007;  
 RA Bowen M.A., Patel D.D., Li X., Modrell B., Malacko A.R.,  
 RA Wang W.-C., Margard H., Neubauer M., Pesano J.M., Francke U.,  
 RA Haynes B.F., Aruffo A.;  
 RT "Cloning, mapping, and characterization of activated leukocyte-cell  
 RT adhesion molecule (ALCAM), a CD6 ligand.";  
 RL J. Exp. Med. 181:2213-2220(1995).  
 RN [2]  
 RP SEQUENCE OF 2-583 FROM N.A.  
 RX MEDLINE=98161527; PubMed=9502422;  
 RA Degen W.G., van Kempen L.C., Gijzen E.G., van Groningen J.J.,  
 RA van Kooyk Y., Bloemers H.P., Smit G.W.,  
 RT "MEM, a new cell adhesion molecule in metastasizing human melanoma  
 RT cell lines, is identical to ALCAM (activated leukocyte cell adhesion  
 RT molecule).";  
 RL Am. J. Pathol. 152:805-813(1998).

Query Match	Similarity	14.6%	Score 492.5	DB 1:	Length 583:
Best Local	Similarity	24.8%	Pred. No. 2.4e-25:		
Matches 145:	Conservative 121:	Mismatches 256:	Indels 63:	Gaps 23:	
QY	41	GSTALLKCGLSQSGNLSHVDWFSYKEREKTLTFRRQGGOGS---	EPGEYQRSLSD	96	
Db	36	GDIIIFPCRLDVPQ-NLMFGKMKTEKRPDQSPFIARSRSTKKSQYDDVPEYKDRNLSE	94		
QY	97	RGATLLAQVTPDDEIFLCQ-GKRRSQEYRIQLRVYAPAEPEPINQVPLGIPVNSKP	155		
Db	95	-NTTISISARISDEKRFECMLVTEEDNVEAEPTIYKVFQPKSPETISKALFL-ETEO	151		
QY	156	EEVATVGNNGPYIPQIVYKNGRPLKEKKNVHIQSQTWS-SGLYTLQSLKQALYK	214		
Db	152	KRIQGDQISDSYDPDGNITMYRNGKVLHPLEGAVIIIFKEMQPVYQITMTSTLEKTKT	211		
QY	215	EDKDAQFYELNRLPFGNNHKESEEVYVPEYPEKWLVEY-EPVGMLEKGDREYRCL	273		
Db	212	ADIQMFPTCSVYVYFGSGOKTTHSEQAVEDIYPTQEVYIQLVPPKMAIKEDNITLKL	271		
QY	274	ADGNPPH---PSISQNSNSTEAEETITNDQVGLVLEPAKREHSGRKECGSLDPTMIS	330		
Db	272	GNQNPPEEFLYLLPEQPGIRSSNTYTLMD-----VRNNAQGVKCSLIDKKSMTA	323		
QY	331	LSEPEQLLVNVSDVRSAPAP-ERQEGSSLTITCEASSODLEFQMLRETEQVLEBGR	389		
Db	324	STP---ITVHTL-DLSLNPGEVTRQIGDALPVSCITISASNNALVYMKNDIR--LSS	376		
QY	390	PVQLQHLDKREKGGGKCYAVSPSTPGNLRIQLVNVALFEPMPMAKERRKYVKNVNLN	449		
Db	377	P--SFSSLTMYQACNVYVETALQAEVGEKLRKRETLIVEGKP-QIKMKTKTDPSGLSKT	432		
QY	450	LSCGASGHPRPITSMNVNGTAS---EQODDP---RVSLTNLVLTPELLETGVCTASN	503		
Db	433	ICCHGEGFPKPAIQMTITIGSGVINQTEESPINGRYIS-KIITSPENAVT-LTCAEN	489		
QY	504	DIGKNTSLILELVNLTLLTPDSNTTGLSTSTASPHTRANSTERRKLPEPESRGVY	563		
Db	490	QLERTVNSLNSAISI-----PE-----HDEADEISDENR-EKVNDQAKLIV	530		
QY	564	AVTQIILVAVIGAVLYFLY-KKGLPGRSGKQDITLPPSRKSE	607		
Db	531	GVVGLLLAALVAGVYWLKYNKKSKTASKHYNKDGLGNMEENKLE	575		
RESULT	5				
C166_MOUSE		STANDARD:	PRT:	583 AA.	
AC	061490: 070136:				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	CD166 ANTIGEN PRECURSOR (ACTIVATED LEUCOCYTE-CELL ADHESION MOLECULE)				
DE	ADICAM. (DM-GRASP PROTEIN).				
GN	ALCAM.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=NS:				
RX	MEDLINE=97353242; Pubmed=9209500;				
RA	Robert M.A., Bajorath J., D'Egidio M., Whitney G.S., Palmer D.,				
RT	Kobay J., Staring G.C., Sladac A.W., Aruffo A.				
RT	"Characterization of mouse ALCAM (CD166): the CD6 binding domain is				
RL	conserved in different homologs and mediates cross-species binding."				
RL	Eur. J. Immunol. 27:1469-1478(1997).				
RL	[2]				
RP	SEQUENCE OF 227-583 FROM N.A.				
RC	STRAIN=BALB/C; TISSUE=BRAIN;				
RX	MEDLINE=94376084; Pubmed=8089660;				

RA Kanki J.P., Chang S., Kuwada J.Y.:  
 RT "The molecular cloning and characterization of potential chick  
 RT DM-GRASP homologs in zebrafish and mouse."  
 RL J. Neurobiol. 25:831-845(1994).  
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN  
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC  
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO  
 CC ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF  
 CC THE NERVOUS SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U95030; AAC06342.1; -  
 DR MGD: L25274; AAA37528.1; -  
 DR INTERPRO: IPR000485; -  
 DR INTERPRO: IPR003006; -  
 DR PFM: PF00047; 19; 5.  
 DR PROSITE: PS00290; IG-MHC; FALSE\_NEG.  
 DR Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;  
 KW Signal.  
 FT CHAIN 1 27 POTENTIAL.  
 FT SIGNAL 28 583 COI66 ANTIGEN.  
 FT DOMAIN 28 527 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 528 549 POTENTIAL.  
 FT DOMAIN 550 583 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 263 321 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 347 399 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 428 492 IG-LIKE C2-TYPE DOMAIN.  
 FT DISULFID 43 113 POTENTIAL.  
 FT DISULFID 157 220 POTENTIAL.  
 FT DISULFID 270 313 POTENTIAL.  
 FT DISULFID 354 392 POTENTIAL.  
 FT DISULFID 435 485 POTENTIAL.  
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 227 232 PSOKT -> AAGIPA (IN REF. 2).  
 FT CONFLICT 454 454 S -> F (IN REF. 2).  
 FT SEQUENCE 583 AA; 65161 MW; E7BAF8BCA8F9489 CRC64;  
 SQ  
 Query Match 14.5%; Score 488.5; DB 1; Length 583;  
 Best Local Similarity 24.5%; Prd. No. 4.4e-25;  
 Matches 152; Conservative 122; Mismatches 274; Indels 73; Gaps 25;  
 QY 5 RLVCACFLAACCCCPVAGVGEAEQAPAELEVEVSTALLKGLCSQSGNLSHVWFS 64  
 DB 10 RLVECLLISAVALRRLGLWY-----TVNSAYGGTIVMPCLRPDPQ-NLFGKWKY 58  
 QY 65 VHKERTLLFRVROGQGS-----EPGEYERLSLDGKATLALQVPPDERIFLQ-GK 119  
 DB 59 EKPGSPVFIAFRSSTRKSVQYDVDPKRLSLSE-NYTLISINAKISDEKRVCHLYT 117  
 QY 120 RPRSGEYRIQAVKAPPEPNIVPLGIVNSKEPEEVATCGRNGYPIFQVITWNGR 179  
 DB 118 EDNVEAFATLVKVPKSPKPEI-VNKAPE-LETDOQLKLLGICISRDSYPCDNITWYNGK 175  
 QY 180 PLKEKNRVHTIOSOTVE-SSGLYTLQSLIAQLVKEDKDAQFYCELVNRLPSGNHKKES 238

DB 176 VLQVEGEVALLFEKKEIDPQTGLTIVTSSLEKKTTRSDIQMPFTCSVTYVGGPGQRTYS 235  
 QY 239 REVTVPVFPTPEYKALEV-EPVGMLEKGDREYETRCGLADGNPPH---FSTSKONPSTREA 294  
 DB 236 EQEILFDIYPTPEQYIOLVLPKNAIKEGDNTTLOCLNGNPPPEEFYLPQGPGRIRSS 295  
 QY 295 EELTNDNGVLVLPARKHRESGRECOGLDPTITSLSEPOELLVNVSDVRSVPAAP- 353  
 DB 296 NPTLTLD-----VRNATGDKCSLIDKRNMAST---ITVHYL-DLSINPRGEV 342  
 QY 354 EROESSLTITLCEAESODLEFOMLREETGOVLERGEVLOLHDIKRAGGCIYCVSVPS 413  
 DB 343 TKQIGDITLPVSCITISASRNATVVMKDNIR--LRSSP--SFSSHYODAGNVCETALQE 398  
 QY 414 IGPLNRIDLVNVAIFGPPWMAFERKRVKVMKNMVLNLSCEASGPRPTISMNVTAS-- 471  
 DB 399 VEGLEKRRSLTLIVEGKP--QIKMTKTPDPSGLSKTTICHVEGPKPAIMHTTIGSGSVI 456  
 QY 472 -EODDDPQ---RVLSLWLVTPLELTGVECTASNDLGKNTSILFELVNLTPPDSN 527  
 DB 457 NQTESPIYNGRYYS--KIISPEENVY-LTCTANQLEFVNSLVNLSAISI---PE-- 507  
 QY 528 TTTGLSTSPHTRANSTSTERKLPPEPSRGVYIVAVICTVLAVLGAVALTYLY-KKG 586  
 DB 508 -----HDEADDISDENR-EKVNDQAKLIVIGVLLAALVAGVYVWLPMKKS 554  
 QY 587 KLPGRSGKOETITLPPSRKE 607  
 DB 555 KTSKRVNKDGLGNMEKNKLE 575  
 RESULT 6  
 C166\_CARAU STANDARD: PRT: 555 AA.  
 ID C166\_CARAU  
 AC 090304; 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE C166 ANTIGEN HOMOLOG PRECURSOR (NEUROLIN) (DM-GRASP HOMOLOG).  
 OS Carassius auratus (Goldfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=RETINA;  
 RC TISSUE=RETINA;  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=RETINA;  
 RC MEDLINE=94299040; PubMed=8026643;  
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;  
 RT "Molecular characterization of fish neurolin: a growth-associated  
 RT cell surface protein and member of the immunoglobulin superfamily in  
 RT the fish retinotectal system with similarities to chick protein  
 RT DM-GRASP/SC-1/BEN."  
 RL Differentiation 56:21-29(1994).  
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY  
 CC NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCs)  
 CC AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST  
 CC OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED  
 CC FROM THE RGCs AT THE RETINAL MARGIN. REMAINS ON ADULT RGCs ONLY AT  
 CC CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL  
 CC AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: L25056; AAC38015.2; -.

DR HSBP; Q13740; IKJC.

DR INTERPRO: IPR000495; -.

DR PFM; PF00047; 1g; 3.

DR PROSITE: PS00290; IG\_MHC; 1.

KW Signal, Developmental protein; Cell adhesion; Immunoglobulin domain;  
KW Glycoprotein; Transmembrane.

FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 555 CD166 ANTIGEN HOMOLOG.

FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 500 520 POTENTIAL.

FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 31 117 IG-LIKE V-TYPE DOMAIN.

FT DOMAIN 147 224 IG-LIKE V-TYPE DOMAIN.

FT DOMAIN 256 313 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 340 392 IG-LIKE C2-TYPE DOMAIN.

FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN.

FT DISULFID 38 110 POTENTIAL.

FT DISULFID 154 217 POTENTIAL.

FT DISULFID 263 306 POTENTIAL.

FT CARBOHYD 92 92 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 171 171 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 350 350 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 441 441 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

FT CARBOHYD 465 465 N-LINKED (GLCNAC; .) (POTENTIAL).

Db 433 KPDQWVNGNENDESVNNGKATYKLTIVSKUL---VSLVTNKLGEDETEI----- 483

Qy 519 LTTTPDSNTTGTSTASPHRANSTSTERKLPEPSKGVYAVIICILAVLAGAV 578

Db 484 -----SVFSQKNEDGTE-----QAKYIVGIIVGLLVAAALVGL 516

Qy 579 LYFLY-KKGLPKRRSKQEL-TLPSRKSE 607

Db 517 IYWIYIKTRQSGWKGTEKGEAGSESKLE 547

RESULT 7

ID C166\_BRARE STANDARD; PRT; 564 AA.

AC Q90460; Q90480;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CD166 ANTIGEN HOMOLOG PRECURSOR (NEUTROLIN) (DM-GRASP HOMOLOG).

GN CD166.

OS Brachydanio rerio (Zebrafish) (Zebra danio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

OC Cypriniformes; Cyprinidae; Rasbora; Danio.

OC [1]

RP SEQUENCE FROM N.A.

RP MEDLINE-94376084; PubMed-8089660;

Db 433 KPDQWVNGNENDESVNNGKATYKLTIVSKUL---VSLVTNKLGEDETEI----- 483

Qy 519 LTTTPDSNTTGTSTASPHRANSTSTERKLPEPSKGVYAVIICILAVLAGAV 578

Db 484 -----SVFSQKNEDGTE-----QAKYIVGIIVGLLVAAALVGL 516

Qy 579 LYFLY-KKGLPKRRSKQEL-TLPSRKSE 607

Db 517 IYWIYIKTRQSGWKGTEKGEAGSESKLE 547

RESULT 7

ID C166\_BRARE STANDARD; PRT; 564 AA.

AC Q90460; Q90480;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE CD166 ANTIGEN HOMOLOG PRECURSOR (NEUTROLIN) (DM-GRASP HOMOLOG).

GN CD166.

OS Brachydanio rerio (Zebrafish) (Zebra danio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;

OC Cypriniformes; Cyprinidae; Rasbora; Danio.

OC [1]

RP SEQUENCE FROM N.A.

RP MEDLINE-94376084; PubMed-8089660;

```

FT DOMAIN 340 389 IG-LIKE C2-TYPE DOMAIN
FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN
FT DISULFID 38 110 POTENTIAL
FT DISULFID 154 217 POTENTIAL
FT DISULFID 263 306 POTENTIAL
FT DISULFID 426 470 POTENTIAL
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 488 497 LEEDKPKPG -> R (IN REF. 2)
SQ SEQUENCE 564 AA: 61273 MW: ADAA7C293A607929 CRC64;

Query Match 13.0%; Score 438; DB 1; Length 564;
Best Local Similarity 26.0%; Pred. No. 9.2e-22;
Matches 161; Conservative 112; Mismatches 254; Indels 92; Gaps 26;

QY 7 VCAFLAACCCPRVAGPEAQPAPELVEVEVSGTALCKGLSOSQGNLSHVDFSVH 66
DB 12 IAAALFAPGSCLPVIGLYGET-----IEVPCNN-----GNNKPDG-LITTKKRYAK 57
QY 67 KEKRTLLFRVROGOGQ----SEPEGEQRLSLDRGATLALTVYTPDERIFLCQ-GKRP 121
DB 58 DDGSPGDLLIKQAOQKDDPTVSAMDGKTRVISA-ANSSLLIAOGSLTDQGFVTCMVVSSP 116
QY 122 RSQERYRLQVYKAPPEPNQVNPGLGIPVNSKEPEEATCGVNGNGPIQVYWKNGRPL 181
DB 117 NLEEFSEVAVYHKKPAPVIR-KNVKLENGK-LTQLEGCVESANPADLIMKKNQAL 174
QY 182 KEENRHHIOSQTVES-SGLYTLQSLILKAQLVKEKDQAFYCELYNLRPSGNHMSRE 240
DB 175 VDQCKTIIITSDVTKDPVTLGISTSSLSLQYTKAREDAVASQFTC-----VAKHVTCPNQ 227
QY 241 VYVP-----VYPEKVMLEVEPNGLKEGDRVETRLADGNPP--HESISKQNSTRE 293
DB 228 VSTPDFQIRPYREKVSILQVVSQSPIREGDDVTLCKQADBNPPTSFNFI-----K 279
QY 294 AEEETNDNGVLELPARKHESGRYEGEGDLDFTMISLSEPOELLVNVYSDVRSFAP 353
DB 280 GKRYVTYDKDYVTLTGVTNRADSGVYKCSLNDNDVMS-----TQIVYVSFL-DASLTPYTK 334
QY 354 E-ROEGSSLLTCEAESODLEFQMLRETFQVLEGRPVQLDLHDKRAGGVCVAVSP 412
DB 335 VLKRLGELVLYSLEKNASSEKVTWTKD--NRKLDKLP--DESQLRVSADGLVYCDV--- 387
QY 413 SIPLNRTQLVNVAIFGPPMA--FKERKVVAKENMVLNLSCEASGHPRTISNVNGTA 470
DB 388 SIEGIKHSFSELTVEGGPRITGLTKHRSNDGKHV---LTCEAEGSPKPEVOMSVNGTD 444
QY 471 SEQQDDPQVRLSTNLNVLVTPPELTGVECTASNDLGKNTSILFELVNLTLTFPDSWTT 530
DB 445 DETSYVNGKATYKLTLYVPSKLT---VSLVTKNLGFDT-----KQISVSLPEEDKPKR 496
QY 531 GLSTASPHTRANSTERKLPEPESRGVYIVAVICIILAVLAVGLVFLY-KKCKLP 589
DB 497 G-----KNEDGADQAKVIVGVVGLFLAALVGLIWLILYTKTKRQOG 537
QY 590 CRSGKQEI-TLPSRKSE 607
DB 538 SMKTEGETGTSESKLE 556

RESULT 8
ICR.DROME STANDARD; PRT; 764 AA.
AC Q08180;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IRREGULAR CHIASM C-ROUGHEST PROTEIN PRECURSOR (IRREC PROTEIN).
GN RST.
OS Drosophila melanogaster (Fruit fly).

```

```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94102535; PubMed=7503814;
RA Ramos R.G., Igloi G.L., Lichte B., Baumann U., Maier D.,
RA Schneider T., Brandstaetter J.H., Froehlich A., Fischbach K.-F.;
RT "The irregular chiasm C-roughest locus of Drosophila, which affects
RT axonal projections and programmed cell death, encodes a novel
RT immunoglobulin-like protein."
RL Genes Dev. 7:2533-2547(1993).
CC -! FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
CC RETINA.
CC -! SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -! TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
CC -! DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
CC IN LATE LARVAL AND PUPAL STAGES.
CC -! SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z21641; CAN79756.1; -
DR EMBL: L11040; AAA16632.1; -
DR PIR: A49448; A49448.
DR PIR: S34129; S34129.
DR HSSP: P56276; ITRK.
DR FLYBASE: FBgn003285; rst.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00047; 1g: 5.
KW Transmembrane; Immunoglobulin domain; glycoprotein; Signal;
KW Cell adhesion.
FT SIGNAL 1 19
FT CHAIN 20 764
FT DOMAIN 20 533
FT TRANSMEM 534 556
FT DOMAIN 557 764
FT DOMAIN 32 115
FT DOMAIN 137 221
FT DOMAIN 245 261
FT DOMAIN 265 332
FT DOMAIN 353 414
FT DOMAIN 432 515
FT DOMAIN 637 660
FT DOMAIN 211 211
FT CARBOHYD 313 313
FT CARBOHYD 393 393
FT CARBOHYD 400 400
FT CARBOHYD 507 507
SQ SEQUENCE 764 AA: 82947 MW: 262225D2B2A1C181 CRC64;

Query Match 8.0%; Score 270; DB 1; Length 764;
Best Local Similarity 20.7%; Pred. No. 1.8e-10;
Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;

QY 40 VGSTALCKGLSOSQGNLSHV-DMFVSHKEKRTLLFRVROGOGSEPEGEYQRLSLDRG 98
DB 41 VGARVTLPCRVINKQGLTQTKNDPGLGTSRDLGFERAMVGSDEGQY----- 90
QY 99 ATLALTVYTPDERIFLCQK-----RPSQERYRLQVYKAPPEPNQVNPGLGIPVNSK 153
DB 91 SLDIYPMDDDARQCVSPGEGQPAIRSFAGLTVLVPPEAKITQGVIVATADR 149
QY 154 EPEEVATCGVNGNGPIQVYWKYKN-GR-----FLKEKKNVHIOSSQTVESSGL 201

```

```

Db 150 KVE--IECVSAGKPAEITWIDGCVLTIDNIEYVPLEDQR----- 192
QY 202 YLQSLAQLVKEKDKQOFYCELVRLPSGNHMKESREVTVPYPTKEVWLEVEPYGM 261
Db 193 FPAKSVLRLTPKKEHHNHFSCAO--NTADRTYRSAKIREVYKAK--VKVNWGS 246
QY 262 LKEG-----DVEIRCLAGCNPP--HFSISKONSTREAE 295
Db 247 LPGAAGSVGAGGSGVHMKSTGRIVERHSGVRLCEKRAANSDVYKRMFINDEPTIGOK 306
QY 236 EETINDGNVLEPARKEHSGRYECQGLDPTMISLSEPELIVNYSDVRSPAPER 355
Db 307 TE-----MVKNVYRKFDHAIK--EYQNSVKSQSDSETLDSVAPSRQRQSMEA 357
QY 356 QEGSSLTITCEAESQDLEFQWMLRETCQVLERGVYQLDHLKREAGGVCVAVSPSIP 415
Db 358 DVGSVVSLTCEVDSNPDEIYWIQHPSDRVVGTSTNLF--SVSNETAGRYCKANVPCYA 416
QY 416 GLNRTQLVNVAIFGPPMAFKERKVVYKENVVLNLSCEASHPRPT--ISMVNG--TASE 472
Db 417 EISAAVYVYK--GSPAIG--SQTQYGLVGTARIKCHASSVPRARHVSMTENGQELTSE 473
QY 473 QDQD-----PQVLSFLNVLYTPELLETGVECTASNDLCKNTSLILELVNLTITP 524
Db 474 SGHDYSILVDVAPGGVKTLLIRDSQAYHYGKYNTVYDGNVVAELQAKKSVSL-- 531
QY 525 DSNITTGISTASHHTANSTSTERKLPESRGVIVAVTICVLAVGAVLYFLYK 584
Db 532 -----LMTIVG-----ISVAVELVYTLT--VVVY---- 555
QY 585 KGLPCRRSGKQOETLTPSRKSELVEYKSDKLPEMGLLOGSGDRAPDQGEKIDL 644
Db 556 --IKCKKRTK--LPP-----ADVISEHQITKNGVSCALPREDRTSNVSDL 597
QY 645 R 645
Db 598 K 598

RESULT 9
PTK7_CHICK STANDARD; PRT; 1051 AA.
AC Q91046;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TYROSINE-PROTEIN KINASE-LIKE 7 PRECURSOR (KINASE LIKE PROTEIN).
GN PTK7 OR KLG.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC BRAIN;
RA MEDLINE=91271300; PubMed=1711213;
RX Chou Y.-H., Hayman M.J.;
RT "Characterization of a member of the immunoglobulin gene superfamily
RT that possibly represents an additional class of growth factor
RT receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:4897-4901(1991).
CC -1- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC THE CATALYTIC ACTIVITY OF TYROSINE KINASE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW, SPLEEN, BURSAS,
CC THYMUS AND BRAIN. WEAKLY EXPRESSED IN FIBROBLASTS. ALSO EXPRESSED
CC IN EMBRYONIC LIVER.
CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC PROTEIN KINASES. TRK-TYPE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; M63437; AAA48933.1; -.
CC HSSP; P00523; PTK.
CC INTERPRO; IPR000719; -.
CC INTERPRO; IPR001245; -.
CC INTERPRO; IPR003006; -.
CC PRAM; PF00047; 19; 7.
CC PRAM; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
CC Immunoglobulin domain; Repeat.
CC SIGNAL 1 22
CC CHAIN 23 1051
CC DOMAIN 23 685
CC TRANSMEM 686 706
CC DOMAIN 707 1051
CC DOMAIN 33 95
CC DOMAIN 130 194
CC DOMAIN 227 289
CC DOMAIN 319 379
CC DOMAIN 407 469
CC DOMAIN 498 558
CC DOMAIN 587 652
CC DOMAIN 777 1048
CC DISULFID 40 88
CC DISULFID 137 187
CC DISULFID 234 282
CC DISULFID 326 372
CC DISULFID 414 462
CC DISULFID 505 551
CC DISULFID 594 645
CC CARBOHYD 103 103
CC CARBOHYD 202 202
CC CARBOHYD 255 255
CC CARBOHYD 264 264
CC CARBOHYD 444 444
CC CARBOHYD 548 548
CC CARBOHYD 627 627
CC SEQUENCE 1051 AA; 116366 MW; 1752442AAECB702 CRC64;

Query Match 7.7%; Score 259.5; DB 1; Length 1051;
Best Local Similarity 20.1%; Pred. No. 1.4e-09;
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

QY 29 EQPAELVEVYSTALLKCGLSQSGNLSHYD-----WF-----SVHK 67
Db 119 KQPA-SAAIQQSSIVYLC-----HIDGPRPTWQFRGADLPDGRGTYSVS 167
QY 68 EKRTLIFRVGGGSGSEGEY-----BQRLSDRGATLALQ-----V 106
Db 168 KERTILTR--GAGPDDNMLYYCSARPAVGSVCQDNFTINIDESPQAVVPEDLIV 224
QY 107 TPQDERIFLCO--GKRPSQOEY-----RQLR--YK-- 134
Db 225 TNERAMFCCQAAVPPPTQEWLFEDSPITNSKTTVEANGSLILITQKARSTGVYCKIG 284
QY 135 -----APEENIQVPLGIPYNSKEPEVATCGRNGYPIPOV 172
Db 285 HGQKQKALVLTALRLAIEEMAPSPKVLNANGHR-----SCAPRGVPTPQV 335
QY 173 IYKKN-----GRLKEKNRVHIQSSQYVSSGLYTLQSLAQLVKEKDKQAFYCEL 225
Db 336 WNRQNERVPTAGRYQEAEDLVF--TSITADAGIYCHAAANKA----- 378

```

QY 226 NRYLPSGNHMKRESREVTVPVFEKVMLEVEPVGMLEKEDRVEIKCLADGNPPPHFSIS 285  
 DB 379 -----GKKOELITV-----ATVPKVMKPKDSOLESKCYLHCLSKASIKPVTMY 427  
 QY 286 KNPSTREAEEETNDNGVLYLEPARKHSGRYECOGDLDTMISLSPQELLVNY 342  
 DB 428 RNCVTSJSDSRFISNGTLRINNVEYDGTMYK-----VSPSPACISIEGYAHV 477  
 QY 343 --VSDRVSPAPBERQ---EGSSLITTCASQDLEFQMLREE---IGCVLERGPVQ 393  
 DB 478 HYLEKIKFTPPQPLQCMENKKEVYSCATGREKPTIQTMTDSSLPSSHRAGLIS 537  
 \* QY 394 LHLKREAGGRCVAVSPSIPGLNRTOLVNAIFGPPMAFK--ERKVVKEMVNLNIS 451  
 DB 538 FHKVYSDSDSNYTCIASNSQGEIRATVOLVAVY---VTFLEPEPTTVQGHAMFQ 593  
 QY 452 CEASGHPRTISMVNGTASEODOPRVLTSLNVLTPELLETEGYECTASNDLKNISI 511  
 DB 594 CQAGDPVPHIQMK-----GKDKILDPKSLPIRIQIMPNSLV---LYDVTEDSGKYTCI 646  
 QY 512 -----LLELVNLTTLTPDSNTTGLSTASPTTRANSSTERKILPEPESRGV 560  
 DB 647 AGNSCNKIKHREALYVVDKPAEEDG-----PSSHTPYKMTQT-----IGL 688  
 QY 561 VIVAVICILVAVLGAVALFYKKGKLPGRSGKQEIPLPGRKSEL----- 608  
 DB 689 SVGAAYITII--VLG--LMFYCK-----RRAKRLKHKHBEDEPEMECNGTILONG 739  
 QY 609 --VVEVSKDLPEEMGLQSS 628  
 DB 740 OTTAEIOEVALTNIGSSSGAS 761  
 RESULT 10  
 NCAL\_XENLA STANDARD: PRT: 1088 AA.  
 ID NCAL\_XENLA  
 AC P16170:  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE NEURAL CELL ADHESION MOLECULE 1, LARGE ISOFORM PRECURSOR (N-CAM 180)  
 [CONTAINS: N-CAM 140].  
 GN NCAM1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodidae; Xenopus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90098871; Pubmed=2481269;  
 RA Krieg P.A., Sakaguchi D.S., Kintner C.R.;  
 RT "Primary structure and developmental expression of a large  
 cytoplasmic domain form of xenopus laevis neural cell adhesion  
 molecule (NCAM).";  
 RL Nucleic Acids Res. 17:10321-10335(1989).  
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF  
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURON AND IN PRESUMPTIVE NEURAL  
 CC TISSUE.  
 CC -1- DEVELOPMENTAL STAGE: THE MRNA ENCODING THIS LD-NCAM IS THE MAJOR  
 CC TRANSCRIPT PRESENT IN BOTH MATERNAL RNA AND IN THE EMBRYO DURING  
 CC EARLY NEURAL DEVELOPMENT.  
 CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M25696; AAA49909.1; .  
 DR PIR: S09600; IXLNL.  
 DR HSSP: P56276; IFLK.  
 DR INTERPRO: IPR001777; .  
 DR INTERPRO: IPR003006; .  
 DR PFAM: PF000041; fn3. 2.  
 DR PFAM: PF000047; 1g. 5.  
 DR Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;  
 KW Immunoglobulin domain; Alternative splicing; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1088  
 FT  
 FT DOMAIN 20 705  
 FT TRANSMEM 706 723  
 FT DOMAIN 724 1088  
 FT DOMAIN 34 100  
 FT DOMAIN 129 193  
 FT DOMAIN 225 289  
 FT DOMAIN 316 386  
 FT DOMAIN 413 480  
 FT DOMAIN 512 589  
 FT DOMAIN 618 686  
 FT DOMAIN 149 153  
 FT DOMAIN 158 162  
 FT DISULFD 41 93  
 FT DISULFD 136 186  
 FT DISULFD 232 282  
 FT DISULFD 323 379  
 FT DISULFD 420 473  
 FT CARBOHYD 82 82  
 FT CARBOHYD 219 219  
 FT CARBOHYD 310 310  
 FT CARBOHYD 341 341  
 FT CARBOHYD 417 417  
 FT CARBOHYD 443 443  
 FT CARBOHYD 472 472  
 FT VARSPIC 804 1049  
 FT SEQUENCE 1088 AA; 117778 MW; 62738B5B03F3E83 CRC64;  
 Query Match 7.6%; Score 256; DB 1; Length 1088;  
 Best Local Similarity 18.1%; Pred. No. 2.5e-09;  
 Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;  
 QY 33 PELVEVEVSTALNCGLSQSGNSLHVDMSVHKEKTLIFRVAGGQSGSPGEYQRL 92  
 DB 26 PDGETSLSESKFFLC--QVSGEARDISWISPTGEKLVY-----QQOI 66  
 QY 93 SL---ODRGATLALNQTVPQDERIFLC-----QGR-----RPRSGEY 126  
 DB 67 SVYRSDYITSLTIYASSQDAGIYKCVASNEAGSECTVNLKLYOKLTFKNAFTPOEF 126  
 QY 127 R-----  
 DB 127 KEGEDAVIICDVSSIPSITIRHKGKDVIFKKDVRFYVLANNYQIKIKTIDGTGTRC 186  
 QY 139 -----  
 DB 187 EGRILARGELNKIDIOYIYVNPPTIOAROLRNANANMAESVLSGDDAGFDPDSMLK 246  
 QY 177 NGRPLKEKNRVHIQSQTVESGLYTLQSLKALQYKEDKDAQFYCELNRYLPSCNIMK 236  
 DB 247 KGEPIEDGEEKISFND-----QSEMTIHVHEQDEDAVEYSCIAN-----NQAG 289  
 QY 237 ESREVEVPPVFPPEKVMLEVEPVGMLEKEDRVEIKCLADGNPPPHFSISKNPSTREAEE 296  
 DB 290 EAEATILLKYAKPKTYVENKTAIV--ELDEITLLTCEASGDPLP--STWRTAVRNISSE 345  
 QY 297 ETTNDNGVLYLEPARKE-----HSGRYEC-----QGLDLDTMISLSLSPQELLVNY 341

```

Db 346 ATTLGDIHIVKEHIRMSALTLDKIDYTDAGEFYCIASNPICVMOAMY-----FEVQ 397
QY 342 YSDVRVSPAPRQESSSTLFCCEASSQDLEFQMLREFTGOVLE-----RGVY 391
Db 398 YAKIR-GPVVYVTTWECNPVNTICEVFAHRAAVTFRD--GOLPSSNFSNIKISGPT 454
QY 392 ---LQHLDLKREAGGRCVY-----ASVPSIPGLNRTOLVN--VAIFG 429
Db 455 SSSLEVVPDSENGNNTCAINTIGHESEFLLVQADPPSSAIKKEVYSSITWAVLD 514
QY 430 P-----WMAFKERKVVVK-----ENM--VLNLSCEAS----- 455
Db 515 EPDSTGVPLIKYKAERVIGHEKMTKYYDAKEVNAESIITVMGLKPTSYVWKLAMN 574
QY 456 ---GHRPITSNVN-----GTASE-----QDDPQVR-----LST 483
Db 575 GKGLDSTPQOEFTTQVPEPSAPKLVLGHLSDEGNSIKVDILKQDDGSPIRHYLVNFA 634
QY 484 LNVLV-TPEL-----LETGVE---CTASNDLGKNTSL--FLELVNLTTLT 523
Db 635 LNLMEKPEKRVPSNSHHVWLKALENNVDVYVAENOGKSKPALLSFRTAKTTAT 694
QY 524 PDSNTTGLSTASAPRTANSTSTERKLPESRGVIVAVICILVLAIGAVLYFLY 583
Db 695 ATASAGTGLGTG-----RRSGKEITLPPSRKSELVVEKSKLPREMKLLG 626
QY 584 KKGLKPC-----RRSGKEITLPPSRKSELVVEKSKLPREMKLLG 626
Db 732 KGLGLMCIANPCGKAGPEKAGKIDIEGKAFFSKDESK--EPLEVEVTE----- 778
QY 627 SSGDKRAPGDGGERYID 643
Db 779 ---EERTPNHDSNQIE 792

```

```

RESULT 11
NCAM_MOUSE
ID NCAM_MOUSE STANDARD: PRT; 725 AA.
AC P13594; 061950;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE, PHOSPHATIDYLINOSITOL-LINKED ISOFORM
DE PRECURSOR (N-CAM 120) (NCAM-120).
GN NCAM1 OR NCAM.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-87246524; Pubmed-3555563;
RA Bartheis D., Santoni M.-O., Wille W., Ruppert C., Caix J.-C.,
RA Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region.",
RL EMBO J. 6:907-914(1987).
RN [2]
RP SEQUENCE OF 20-700 FROM N.A.
RC STRAIN-C57BL/6; TISSUE=BRIN;
RX MEDLINE-89251563; Pubmed-2721486;
RA Santoni M.J., Bartheis D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain.",
RL EMBO J. 8:385-392(1989).
RN [3]
RP SEQUENCE OF 642-725 FROM N.A.
RX MEDLINE-88283628; Pubmed-3396534;
RA Barbs J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse.",
RL EMBO J. 7:625-632(1988).

```

```

RN [4]
RP SEQUENCE OF 20-36.
RX MEDLINE-86140120; Pubmed-3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules.",
RL J. Biol. Chem. 261:3396-3401(1986).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00051; CAA68263.1; -
DR EMBL; X15049; CAA33148.1; ALT_SEQ.
DR EMBL; X07195; CAA30173.1; -
DR PIR; A29673; IJMSNG.
DR MGD; MGI:97281; NCAM.
DR INTERPRO; IPR001777; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; ig; 5.
KM Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KM Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT FT SIGNAL 1 19
FT CHAIN 1 20 725
FT FT DOMAIN 34 103
FT FT DOMAIN 132 196
FT FT DOMAIN 228 295
FT FT DOMAIN 323 393
FT FT DOMAIN 420 487
FT FT DOMAIN 519 596
FT FT DOMAIN 625 692
FT FT DOMAIN 152 156
FT FT DOMAIN 161 165
FT FT DISULFID 41 96
FT FT DISULFID 139 189
FT FT DISULFID 235 288
FT FT DISULFID 330 386
FT FT DISULFID 427 480
FT FT CARBOHYD 222 222
FT FT CARBOHYD 316 316
FT FT CARBOHYD 348 348
FT FT CARBOHYD 424 424
FT FT CARBOHYD 450 450
FT FT CARBOHYD 479 479
FT FT CARBOHYD 261 268
FT FT CONFLICT 273 273
FT FT CONFLICT 354 355
FT FT CONFLICT 549 549
FT FT CONFLICT 572 572
FT FT CONFLICT 575 575
FT FT CONFLICT 589 594
FT FT CONFLICT 600 602
FT FT CONFLICT 657 657
SQ SEQUENCE 725 AA; 80296 MW; C2AEBB84461C6B2F CRC64;

```

Query Match 7.3%; Score 246.5; DB 1; Length 725;  
 Best Local Similarity 20.6%; Pred. No. 6.1e-09;  
 Matches 141; Conservative 92; Mismatches 261; Indels 189; Gaps 33;





```

OY 553 PEPESRGVIVAV-----IVCLTVL-----AVLGAVL 579
DB 531 DEPEATGCVPIIKYAEKMSLGEAMHSHKWTDAKANEGLVTTIMGLPEPTRIYALAL 590
OY 580 YFLYKKGKLPCCRRSGKQITLPPSKSELVYEVKSDKLPBEMG 622
DB 591 -----NGK-----GLGEISATATEKFTQPVREPSAPKLEGQMG 622

RESULT 14
PCBM_MOUSE
ID PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
GN HSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=92078153; PubMed=1744087;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogell G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RA "The complete sequence of perlecan, a basement membrane heparan
RT sulfate proteoglycan, reveals extensive similarity with laminin A
RT chain, low density lipoprotein-receptor, and the neural cell adhesion
RT molecule."
RL J. Biol. Chem. 266:22939-22947(1991).
RN [2]
RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=89034110; PubMed=2972708;
RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogell G., Sasaki M.,
RA Yamada Y., Hassell J.R.;
RT "Identification of cDNA clones encoding different domains of the
RT basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 263:16379-16387(1988).
CC -1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC SUBSTRATE FOR CELLS.
CC -1- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC DIMERS OR STELLATE STRUCTURES.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 10-5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M77174; AAA39911.1; -
DR EMBL: J04054; AAA39899.1; -
DR EMBL: J04055; AAA39912.1; -
DR HSPG; P01130; IAUJ.

```

```

DR MGD: MGI:96257; HSPG2.
DR INTERPRO: IPR000034; -
DR INTERPRO: IPR000082; -
DR INTERPRO: IPR000561; -
DR INTERPRO: IPR001438; -
DR INTERPRO: IPR001791; -
DR INTERPRO: IPR002049; -
DR INTERPRO: IPR002172; -
DR INTERPRO: IPR003006; -
DR PFAM: PF001390; SEA; 1.
DR PFAM: PF00047; 19; 15.
DR PFAM: PF00052; laminin_B; 3.
DR PFAM: PF00053; laminin_EGF; 8.
DR PFAM: PF00054; laminin_G; 3.
DR PFAM: PF00057; LDL_receptor_a; 4.
DR PRINTS: PR00010; EGFBLD.
DR PROSITE: PS00022; EGF_1; 8.
DR PROSITE: PS01186; EGF_2; 5.
DR PROSITE: PS01209; LDLRA_1; 4.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
DR PROSITE: PS00068; LDLRA_2; 4.
KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW Heparan sulfate; laminin EGF-like domain; Immunoglobulin domain;
KW Extracellular matrix; EGF-like domain.
FT CHAIN 1 21
FT 22 3707
FT 22 193
FT 194 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 764 813
FT 814 871
FT 879 923
FT 924 933
FT 934 1125
FT 1126 1158
FT 1159 1208
FT 1209 1265
FT 1275 1324
FT 1325 1334
FT 1335 1529
FT 1530 1562
FT 1563 1612
FT 1613 1670
FT 1677 1771
FT 1772 1865
FT 1866 1954
FT 1955 2049
FT 2050 2148
FT 2149 2244
FT 2245 2343
FT 2437 2436
FT 2533 2619
FT 2620 2720
FT 2721 2809
FT 2810 2895
FT 2896 2980
FT 2980 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 507 1676
FT 1677 2980
FT 2981 3707
FT 194 234
FT 281 319
FT 320 359
FT 360 403
FT 404 504
FT 521 530
FT 531 730
FT 731 763
FT 7
```



```

FT DOMAIN 2961 3130 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT DOMAIN 3049 3241 EGF-LIKE 1.
FT DOMAIN 3304 3495 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT SITE 3558 3705 HEPARAN SULFATE (POTENTIAL).
FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 HEPARAN SULFATE (POTENTIAL).
FT SITE 3615 3617 MEDIATES MOTOR NEURON ATTACHMENT (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 428 479 BY SIMILARITY.
FT DISULFID 764 773 BY SIMILARITY.
FT DISULFID 766 780 BY SIMILARITY.
FT DISULFID 783 792 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 814 829 BY SIMILARITY.
FT DISULFID 816 839 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 869 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
FT DISULFID 1161 1175 BY SIMILARITY.
FT DISULFID 1178 1187 BY SIMILARITY.
FT DISULFID 1190 1206 BY SIMILARITY.
FT DISULFID 1209 1224 BY SIMILARITY.
FT DISULFID 1211 1234 BY SIMILARITY.
FT DISULFID 1237 1246 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1275 1287 BY SIMILARITY.
FT DISULFID 1277 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1307 1322 BY SIMILARITY.
FT DISULFID 1363 1572 BY SIMILARITY.
FT DISULFID 1565 1579 BY SIMILARITY.
FT DISULFID 1582 1591 BY SIMILARITY.
FT DISULFID 1594 1610 BY SIMILARITY.
FT DISULFID 1613 1628 BY SIMILARITY.
FT DISULFID 1615 1638 BY SIMILARITY.
FT DISULFID 1641 1650 BY SIMILARITY.
FT DISULFID 1653 1668 BY SIMILARITY.
FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.
FT DISULFID 2365 2413 BY SIMILARITY.
FT DISULFID 2456 2506 BY SIMILARITY.
FT DISULFID 2534 2599 BY SIMILARITY.
FT DISULFID 2641 2686 BY SIMILARITY.
FT DISULFID 2831 2876 BY SIMILARITY.
FT DISULFID 2917 2962 BY SIMILARITY.
FT CARBOHYD 65 65 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 71 71 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 76 76 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).

```

Query Match 7.3%; Score 244.5; DB 1; Length 3707;

Best Local Similarity 22.3%; Pred No. 8e-08; Matches 113; Conservative 79; Mismatches 190; Indels 125; Gaps 26;

QY 17 CCRPVAGVP---GEAEQAPAELEVEV-----GSTALIKCKGLSOGNLSHVDF 63

```

Db 1653 CAPGEGDPNAGGRCQPLTESLEVOIHBSRSVPOGGGHSILRC---QVSGSPPHYFW 1709
QY 64 SVHKERKTLIFRVQOGQSEGEYEORLSLQDGATLALTQVTPQDERFLFGQGR-PR 122
Db 1710 S-----REDGRPLPSSAQOR---HGSLEHPPSVQSPAGYITCTCRLLH 1752
QY 123 SQEIRQLRYKAPDEEPNIOVPLGIVNSKEPEVA-----TCVRNGYDIPQVIW 175
Db 1753 TNSRAELVAEAPSK-----PIWVTEQRSQSVAPGADVFTICAKSPAYILWMT 1806
QY 176 K-NGRPLKEKKNVHIIQSSQTVSSGLTLOSLTKQVLEKDKAQFY-CEIN--YRLP 230
Db 1807 RLHNGK-----LPSRADFNGILTRVY-----QPSDGLTYVCTSSNMFAMD 1848
QY 231 SGN---HMKSEREVTVV--FYPTKEVLEVEPVGMLEKEDRAVEICLADGNPPHF--- 282
Db 1849 OGATLHVQVSGSTAVASIHPPQ--LVQP-----GQAFKCSATGNTPTMLEMT 1899
QY 283 -SISKQNPSTREAEETTNQGVVLEPARKHSGRYECQGLD-----LDTMISLL--S 333
Db 1900 GGPSGQLPAAQIH-----NGILRLPALEPSPDQGYLCRLSSAGHVARAMLQVHGGS 1953
QY 334 EPQELVNVYSDVRSVPAPEQSGSLTLTCEASSQDLEFQWLAETGVLER---G 389
Db 1954 GPR-----VOVSPEKQVHEGRTVRLCYCAAGVPSASITW--RKEGGSLEPRQAHG 2003
QY 390 PVLQHLDKREAGGRCVAVSPSIPGLNFTOLVY-----ATGPPMAFKERKVV 441
Db 2004 SRLRLHMSVADSGEYVCGRAN--NNIDAQETSIMISVSPNSPPAPASAPRIRESSSR 2062
QY 442 VKENMVNTLSCSASGHPRTISNVNG 468
Db 2063 VAEQTLIDLNCVVGHAHQVTHKRG 2089

RESULT 15
NCAL BOVIN STANDARD; PRT: 853 AA.
AC P31836;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE, 140 KDA ISOFORM PRECURSOR (N-CAM 140) (NCAM-140).
GN NCAM OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-BRAIN CORTEX.
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakiлина T.V., Feshchenko E.A., Ishchenko K.A.,
RA Mirzoeva S.F., Chernova M.N., Dranysheva S.M.;
RT "Calmodulin-independent bovine brain adenylate cyclase. Amino acid
RL sequence and nucleotide sequence of the corresponding cDNA.";
RL FEBS Lett. 251:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RL domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;
RT "A bovine brain CDNA purported to encode calmodulin-insensitive
RL adenylate cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:42 ; Search time 36.94 Seconds  
(without alignments)  
2049.711 Million cell updates/sec

Title: US-09-653-961-2  
Perfect score: 3363  
Sequence: 1 MGLPRIVCAFLIACCCCP...SSGDKRAPGQGEKYIDLKH 646

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPTRMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhnc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3356	99.8	646	4	095812 homo sapien
2	1104	32.8	626	13	098922 gallus gall
3	1025	30.5	626	13	090880 gallus gall
4	936	27.8	584	13	098921 gallus gall
5	863	25.7	584	13	090989 gallus gall
6	747	22.2	504	13	096923 gallus gall
7	618	18.4	622	11	09JKB2 mus musculu
8	614	18.3	622	11	09R069 mus musculu
9	499.5	14.9	583	11	035112 rattus norv
10	485	14.4	521	6	046634 canis fami
11	465	13.8	521	6	046631 oryctolagu
12	407	12.1	88	11	09JH02 mus musculu
13	362	10.8	86	6	019129 bos tauru
14	361	10.7	355	6	09M208 bos tauru
15	281	8.4	5198	5	076518 caenorhabd
16	268.5	8.0	538	6	029123 sus scrofa
17	264.5	7.9	538	6	028939 sus scrofa
18	248.5	7.4	846	13	057577 cynops pyrr
19	248.5	7.4	1100	13	057576 cynops pyrr

20	247	7.3	1612	11	089026 mus musculu
21	246.5	7.3	1651	11	055005 rattus norv
22	245.5	7.3	739	11	063669 rattus norv
23	244.5	7.3	823	5	09V010 drosophila
24	242	7.2	739	6	028260 canis fami
25	240	7.1	1482	5	09V4Y0 drosophila
26	237	7.0	725	13	073633 xenopus lae
27	236	6.9	7962	4	010465 homo sapien
28	233	6.9	1496	4	092626 homo sapien
29	231.5	6.9	1259	11	090X38 mus musculu
30	231	6.9	734	4	09N059 mus musculu
31	231	6.9	1344	11	092214 mus musculu
32	230.5	6.9	968	5	09W4Y9 drosophila
33	230	6.8	4351	5	09W4Y4 drosophila
34	229.5	6.8	761	5	09N9Y9 drosophila
35	229.5	6.8	959	5	09NCP3 drosophila
36	229	6.8	734	4	09NPG9 drosophila
37	228.5	6.8	534	5	09VSG5 drosophila
38	228.5	6.8	734	4	09N0Y0 drosophila
39	228.5	6.8	734	4	09NPG5 drosophila
40	228	6.8	815	5	09W4M6 drosophila
41	227	6.7	1026	4	094780 homo sapien
42	227	6.7	1100	4	094779 homo sapien
43	226	6.7	734	4	09NQT2 drosophila
44	226	6.7	734	4	09NPG7 drosophila
45	225.5	6.7	734	4	09NQT1 drosophila

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	646 AA
095812	095812			
AC	095812			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)			
DE	CELL SURFACE GLYCOPROTEIN P1H2 PRECURSOR.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Gui L., Chang L., Browne P.V., Hebbel R.P.:			
RT	"P1H2 from human umbilical vein endothelial cells."			
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF089868; AAD17799.1; -			
DR	INTERPRO: IPR003006; -			
DR	PFAM: PF00047; 1g; 5.			
DR	SEQUENCE 646 AA; 71665 MW; 1B5FB8ADD930738E CRC64;			
SO				
Query Match	Best Local Similarity	99.8%;	Score 3356;	DB 4; Length 646;
Matches 645;	Conservative	0;	Mismatches 1;	Indels 0;
0;				
QY	1	MGLPRIVCAFLIACCCCPRYAGVGEAEQAPAPLVEVEVSTALIKGLSOSQGNLSHV	60	
DB	1	MGLPRIVCAFLIACCCCPRYAGVGEAEQAPAPLVEVEVSTALIKGLSOSQGNLSHV	60	
QY	61	DMFSVHKRRTLLRRVVGQSGSEGEYRSLSDRATATLQVPPDRIFLCQKR	120	
DB	61	DMFSVHKRRTLLRRVVGQSGSEGEYRSLSDRATATLQVPPDRIFLCQKR	120	
QY	61	DMFSVHKRRTLLRRVVGQSGSEGEYRSLSDRATATLQVPPDRIFLCQKR	120	
DB	61	DMFSVHKRRTLLRRVVGQSGSEGEYRSLSDRATATLQVPPDRIFLCQKR	120	
QY	121	PRSEYRIQLRVYKAPPEPNIOVNPDLGIPVNSKEPEEATCGNGYPIPOVIWYKNRP	180	
DB	121	PRSEYRIQLRVYKAPPEPNIOVNPDLGIPVNSKEPEEATCGNGYPIPOVIWYKNRP	180	
QY	181	LKEKNRVHIQSSQTVESGLYTIQSILKAQLVKEDKDAQFYCELINRLPSGNHMKESRE	240	
DB	181	LKEKNRVHIQSSQTVESGLYTIQSILKAQLVKEDKDAQFYCELINRLPSGNHMKESRE	240	

```

Oy 241 VTVPEVPTKEKWLVEPEVGMLEKGRVREICRLADGNPPHFSISKONPSTREAEEETT 300
    |||||
Db 241 VTVPEVPTKEKWLVEPEVGMLEKGRVREICRLADGNPPHFSISKONPSTREAEEETT 300
Oy 301 DNGVLVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVSDVRSVPAPEROEGSS 360
    |||||
Db 301 DNGVLVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVSDVRSVPAPEROEGSS 360
Oy 361 LTTTGEAASSODLEFQWMLREETQVLERGVYLQHLDLKREAGGGRCAVASPSIPGLNRT 420
    |||||
Db 361 LTTTGEAASSODLEFQWMLREETQVLERGVYLQHLDLKREAGGGRCAVASPSIPGLNRT 420
Oy 421 QLVNVAIGPPMAFKEKRWKENVNLNLSCEASGHPRTISMVNVNSTASEODODPORV 480
    |||||
Db 421 QLVNVAIGPPMAFKEKRWKENVNLNLSCEASGHPRTISMVNVNSTASEODODPORV 480
Oy 481 LSTLVNLVTPPELLETGVECTASNDLGKNTSILFELVNLTTLPDSTNTTGLSTASPH 540
    |||||
Db 481 LSTLVNLVTPPELLETGVECTASNDLGKNTSILFELVNLTTLPDSTNTTGLSTASPH 540
Oy 541 TRANTSTERTKLPESRGVTVAVICILVAVLGAVLYLTKGKLPGRSGKQETTL 600
    |||||
Db 541 TRANTSTERTKLPESRGVTVAVICILVAVLGAVLYLTKGKLPGRSGKQETTL 600
Oy 601 PPSRKSELVVEYKSDKLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646
    |||||
Db 601 PPSRKSELVVEYKSDKLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646

```

```

RESULT 2
O98922 PRELIMINARY: PRT: 626 AA.
ID 098922
AC 098922
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H-B190V+; TISSUE=THYMUS;
RX MEDLINE=9713433; PubMed=8978830;
RA Valin O., Dunon D., Aissel F., Dangy J.P., McNaghy K.M., Imhof B.A.;
RT "HEMCAm, an adhesion molecule expressed by c-kit+ hemopoietic
RT progenitors.";
DR J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08855; CAA70080.1;
DR INTERPRO: IPR003006;
DR PFAM: PF00047; 19; 5.
KW Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 626 POTENTIAL.
SQ SEQUENCE 626 AA: 69104 MW: 278F2F47044E163 CRC64;

```

Query Match 32.8%; Score 1104; DB 13; Length 626;  
 Best Local Similarity 38.3%; Pred. No. 6.3e-78;  
 Matches 251; Conservative 116; Mismatches 221; Indels 68; Gaps 16;

```

Oy 8 CAFLAACCCCPRVAGVGEAEPAPLVEVEVSTALLKCGLS-QSOGNLSHVDWFSVH 66
    |||||
Db 22 CPEFLCLLCC-----GAAGRLVYMPAVLEVEIGSTARLESFSTPGNASTSIEMFVFN 76
Oy 67 KEKRTLIRFVRQGG-----OSEPGEYRQLSLQDRGATLALTQVTPDDE-RIFLQ 117
    |||||
Db 77 RRP-----LRSGEAVRHHSAGVRIDETFEYSERLSVGEDKA-LTSKTRQDNARTFLCQ 129
Oy 118 --GKRPRSOEYRIQLRVYKABEENIYNPLGIPVNSKEPEEVATCVGRNGYPIPIQVIMY 175

```

```

Db 130 VGADSGVGSRRRELTYTKIPAPPEITPNASGIPAGSDMLKIAQTSBNSPSPNITMY 189
    |||||
Oy 176 KNGRPLKEKNRHHIOSSQVVESSGLTYLQSLKADLVEDDAQCYCELNRLPSGNIM 235
    |||||
Db 190 KNGEPLLEEDDKTKILTTLVRESNGLYTVVSTLFSKVTREDRSLPHCVHWLQGMRT 249
Oy 236 KESREYVVPVFEPTKEKWLVEEP-VGMLKEGDEVEIRCLADGNPPHFSISKONPSTR 292
    |||||
Db 250 KDSERVAVTVFPTPEHVELRVAATNAGIVKEGDDVKLVCADGNAPAVESFFRRELDSNQ 309
Oy 293 EAEETTNDNGVLVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVSDVRY--SP 350
    |||||
Db 310 DMTSLADTNDGVLMLHNVSKSSGLYRCQTLDDDTQHEGD-VELVNVYIGVQYKMP 368
Oy 351 AAPEROGSSLTLTCEASQODLEFQWMLREETQVLERGVYLQHLDLKREAGGGRCAVAS 410
    |||||
Db 369 SSP-LHEGDSVRLSCTAHSPVKLDYQW-RDARGRKAEONULLNLNLTETSSNSCRVK 426
Oy 411 VPSIPGLNRTQLVNVVAIFGPPMAFKEKRWKENVNLNLSCEASGHPRTISMVNVNGTA 470
    |||||
Db 427 ARSVPLGEQSKQVAVAVKGRPRIVALSAPLYRQDEVINLTCKALAFQPSFHWMSINGTT 486
Oy 471 SEDQDQFQVLSLTVNLVTPPELLETGVECTASNDLGKNTSILFELVNLTTLPDSTNTT 530
    |||||
Db 487 HEY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHIDLDKPS----- 535
Oy 531 GLSTSTASPTTRANSSTERTKLPESRGVTVAVICILVAVLGAVLYLTKGKLP 590
    |||||
Db 536 -----ESKGIITVAITVLVAVLQSTITFPLHKKKISG 570
Oy 591 RRSQKQETLPPSRKSELVVEYKSDKLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646
    |||||
Db 571 GRSGKQDITKPEARKQKNVVEYKSDKLSEAGALQGANCKEKRSPAQSGEKTYDLRH 626

```

```

RESULT 3
O90880 PRELIMINARY: PRT: 626 AA.
ID O90880
AC O90880
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE L-GICERIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEGHORN; TISSUE=GIZZARD;
RX MEDLINE=96081930; PubMed=749388;
RA Taira E., Nagino T., Tanizura H., Takahata N., Kim C., Kuo C., Li B.,
RA Higuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of glicerin, an
RT immunoglobulin superfamily cell adhesion molecule.";
DR J. Biol. Chem. 270:26861-26867(1995).
DR EMBL: D49849; BAA08648.1;
DR INTERPRO: IPR003006;
DR PFAM: PF00047; 19; 5.
SQ SEQUENCE 626 AA: 69074 MW: C4791EBC2EC55133 CRC64;

```

Query Match 30.5%; Score 1025; DB 13; Length 626;  
 Best Local Similarity 37.4%; Pred. No. 9.4e-72;  
 Matches 248; Conservative 114; Mismatches 219; Indels 82; Gaps 19;

```

Oy 8 CAFLAACCCCPRVAGVGEAEPAPLVEVEVSTALLKCGP-----QSOGNLSHVDM 62
    |||||
Db 22 CPEFLCLLCC-----GAAGR-----EYVPPSSA-----GSGDEHSGAQVQLH-FW 63
Oy 63 FSVHKEKRTLIRFVRQGG-----OSEPGEYRQLSLQDRGATLALTQVTPDDE 111

```

```

Db      64 ECLHLRVRVILCOPRRLBSGAVRHNASGVRIDETFEYSERLSVGEDKA-LSISKVTRDQN 122
Qy      112 -RIFLCQ--GKRPSOEYIOLRYKAPDEEPINOVNPGIPVNSKEPEEVATCVGRNGYR 168
Db      123 ARFICQVADSGVGESESTELTYTKIPAPPEITPNSAIPQOSNDMLKIAOCTSENSP 182
Qy      169 IPQVITVKNGRPLKEKNVHIOSQTVESGTYTLOSILKAOLVKEKDAOFCELYNLR 228
Db      183 SPRTITVKNGEPLLEDEKTKILITLVRESNGLYTVSTLFSKVRREDNSLPHCTVHW 242
Qy      229 LPSGNHAKESREYTVVFEYPTKEVMLEVP--VGMLEKEDRVEIRCLADGNPPPHFSISK 287
Db      243 LQGMRTKQSPRVNVTVPFTEHVELRVATFNAGIVKEGDVLYCQADGNPAVPSFEFR 302
Qy      288 --NPSTREAEETTDNDGVLYLEPARKESGRYECQGLDPTMISLSEPOELVNYVSD 345
Db      303 ELGDSWQDMTSLADTNDGYLMLHNWSSSGLYRCOTLDDMTQHEGD--VELVYNYIEG 361
Qy      346 VRY--SPAAPERQESSLTLTCEAESQDLREFQWLEERTGOVLERGPVQLDHLKREAG 403
Db      362 VOYKMEPSSP-LHEDSVRLSTASHSPVKLDYQW--RDANGRYAEGNOILLTNLTFETSS 419
Qy      404 GYRCVASVSPISPLNTOLVNVNAIFGPPMAFEKRVWKENVNLNLSCEASGHPRTIS 463
Db      420 NFSCRYKARVSPGLEOSKQVAVAKPRIVASAPLYRQDEVINLTCKATAFPQPSFH 479
Qy      464 WNVNGTASQDQDPQVRLSTLVNLYTPELLETGVECTASNDGKNTSILFELVNLTLT 523
Db      480 WSVNGTTHEY--MENOHMASNLTVRVSHDLIRAGAMCRVSNALGVSE----KHQILDOK 533
Qy      524 PDSNTTGLSTASPHTRANSTSTERKLPBPESRGVIVAVYICLVAVLGAVALYLY 583
Db      534 PS-----ESKGIITVAITVCLVAVLGAVALYLY 583
Qy      584 KKGLPCRRSGKQETILPSPKSELVEVYKSDKLPEPMGLQSSGDKRAPGQGEKYID 643
Db      564 KKGIISGGRSGKQDITKPRPKRKNVVEYKSDKLSEAGILQGANNEKSPADQSEKYID 623
Qy      644 LRH 646
Db      624 LRN 626

```

RESULT 4

Q98921 PRELIMINARY; PRT: 584 AA.

AC Q98921; 01-FEB-1997 (TREMblrel. 02, Created)

DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)

DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)

DE HEMCAM PRECURSOR.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID-9031;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H.B19OV+; TISSUE=THYMUS;

RX MEDLINE=97133433; PubMed=8978830;

RA Valino O., Dunon D., Aissi F., Dany J.P., McNagay K.M., Imhof B.A.;

RT "HEMCAm, an adhesion molecule expressed by c-kit+ hemopoietic progenitors";

RL J. Cell Biol. 135:1655-1668(1996).

DR EMBL: Y08856; CAA70081.1; -

DR INTERPRO: IPR003006; -

DR PRAM: PF00047; 1g; 5.

KW Signal.

FT SIGNAL 1 35 POTENTIAL.

FT CHAIN 36 584 POTENTIAL.

SEQUENCE 584 AA; 64422 MW; 5972D946604AF6F3 CRC64;

Query Match 27.8%; Score 936; DB 13; Length 584;

Best Local Similarity 35.9%; Pred. No. 7, 7e-65;

Matches 218; Conservative 109; Mismatches 213; Indels 68; Gaps 16;

```

Qy      8 CAFLAACCCPRVAGVGEADAPAPLVEVEVSTALLCGIS-OSQNLSHVDFSVH 66
Db      22 CPFLCLLCC-----GAAGRLEVMPALVEIEISTARLECSFSPGNASFTSIEMEVY 76
Qy      67 KEKRLIRVAGQGS-----QSEPEYRQRLSDRQATLALPGVYQDE-RIFLCQ 117
Db      77 RRP-----LKSQAVRHNASGVRIDETFEYSERLSVGEDKA-LSISKVTRDQNAFTICQ 129
Qy      118 --GKRPSOEYRIOLRYKAPDEEPINOVNPGIPVNSKEPEEVATCVGRNGYRPIQVY 175
Db      130 VGADSGVGESESTELTYTKIPAPPEITPNSAGIPQOSNDMLKIAOCTSENSFPSPNTW 189
Qy      176 KNGRLPEKRNKRVHIOSQTVESGTYTLOSILKAOLVKEKDAOFCELYNRLPSGNHM 235
Db      190 KNGEPLLEDEKTKILITLVRESNGLYTVSTLFSKVRREDNSLPHCTVHWYLOGQMT 249
Qy      236 KESREYTVPVYPTKEVMLEVP--VGMLEKEDRVEIRCLADGNPPPHFSISK--NPSTR 292
Db      250 KDSRVNVTVPFTEHVELRVATFNAGIVKEGDVLYCQADGNPAVPSFEFRRLGDSWQ 309
Qy      293 PAEEETTDNDGVLYLEPARKESGRYECQGLDPTMISLSEPOELVNYVSDVRY--SP 350
Db      310 DMTSLADTNDGYLMLHNWSSSGLYRCOTLDDMTQHEGD--VELVYNYIEGQVMEP 368
Qy      351 AAPREQSSSLTLTCEAESQDLREFQWLEERTGOVLERGPVQLDHLKREAGGYRCVAS 410
Db      369 SSP-LHEDSVRLSTASHSPVKLDYQW--RDANGRYAEGNOILLTNLTFETSSNFSCRYK 426
Qy      411 VPSIPGLRNTOLVNVNAIFGPPMAFEKRVWKENVNLNLSCEASGHPRTISNNVNGTA 470
Db      427 AASVGLQDSQKQVAVAKPRIVASAPLYRQDEVINLTCKATAFPQPSFHMISNGTT 486
Qy      471 SQDQDQDPQVRLSTLVNLYTPELLETGVECTASNDGKNTSILFELVNLTLTTPDSMTT 530
Db      487 HEY-MENOHMASNLTVRVSHDLIRAGAMCRVSNALGVSE-----KHQILDOKS----- 535
Qy      531 GLSTASPHTRANSTSTERKLPBPESRGVIVAVYICLVAVLGAVALYLYKKGLPC 590
Db      536 -----ESKGIITVAITVCLVAVLGAVALYLYLHKKRGISC 570
Qy      591 RRSQKQET 598
Db      571 GRSQKQDI 578

```

RESULT 5

Q90989 PRELIMINARY; PRT: 584 AA.

AC Q90989; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)

DE S-GICERIN PRECURSOR.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID-9031;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=IEGHORN; TISSUE=GIZZARD;

RX MEDLINE=94213753; PubMed=8161457;

RA Taira E., Takaha N., Tanitara H., Kim C., Miki N.;

RT "Molecular cloning and functional expression of gicerin, a novel cell adhesion molecule that binds to neurite outgrowth factor.";

RL Neuron 12:861-872(1994).

FT [2]

SEQUENCE FROM N.A.

RC STRAIN=LEGHORN; TISSUE=GIZZARD;  
RA MEDLINE=96081930; PubMed=7493388;  
RX Taira E., Nagino T., Tanlura H., Takaha N., Kim C., Kuo C., Li B.,  
RA Higuchi H., Miki N.;  
RT "Expression and functional analysis of a novel isoform of gicerin, an  
RT immunoglobulin superfamily cell adhesion molecule.";  
RU J. Biol. Chem. 270:28681-28687(1995).  
DR EMBL: D38559; BAA07563.1; -;  
DR INTERPRO: IPR003006; -;  
DR PFAM: PF00047; 19; 5;  
KW Cell adhesion; Extracellular matrix.  
SQ SEQUENCE 584 AA; 64378 MW; 876E0C3E920BA92F CRC64;

Query Match 25.7%; Score 863; DB 13; Length 584;  
Best Local Similarity 35.1%; Pred. No. 3.9e-59;  
Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps 19;

QY 8 CAFLAACCCCPVAGVGEAPAPPELVEVEVSTALCKGL-----SOGCNLSHVDV 62  
DB 22 CPEFLCLLC-----GAAGRP-----EYMPSSA-----GSGDREHSGVQLLH-PW 63  
QY 63 FSVHKEKRTLIFR---VRGOG-----QSEPEYEQRLSLDRGATLALQVTPDE 111  
DB 64 ECLHLRRVVLQCPRLKSGEAVRNHNSGVRIDETEYSERLSVGEDKA-LSISKVTRDN 122  
QY 112 -RIFLCQ--GKRPSOEYRIQLRVYKAPKEPNIOVNPGLIPVNSKEPEVATCVGRNGYP 168  
DB 123 ARTFCQVGSOGVSGEERTLEYTKIPAPPEITPNSAGIPQSDNMLKIAQCTSENSEFP 182  
QY 169 IPOVITWNGRPLKEKRVHIOSSQVYESSGLTLOSILKAQLVKEDKQAFCELYNR 228  
DB 183 SPNTITWNGRPLQEEKTKILTTLVRESNGLYTVSTLFKVTREDRNSLFHCTVHW 242  
QY 229 LPSGNHMKREVTVPVPEYPTKYLEVER-VGMLKEDGRVEICLADGNPPHFSISKQ 287  
DB 243 LOGQRTKDSRVNVTVYPTHEVELRVATNAGIVKEDDVKLCDDADGNAPAFSEFRR 302  
QY 288 --NPSTRAEETTNDNGVLLEPARKHSGRYCQGLDITMTLSLSEPOELLVNVSD 345  
DB 303 ELGSDMDMTSLADTNDVLMHNVSKSSGLYRCQTLDDMTQHGED-VELVNVYIEG 361  
QY 346 VRV---SPAAREGSSSLTTCEAESSQDLFEQWLRRETOGVLEGRVQLDHLDRAGG 403  
DB 362 VQVMEKESP--LHEGDSVRLSCTAHSPVKLDYQW-RDARGKVAEGNQLLTNTLTFETSS 419  
QY 404 GYRCVASVPISGLNRTQLVNVAFGPPMAFKERKYVMKENVNLNLSCEASGHPRTIS 463  
DB 420 NFSCRVARSVPLEGSKQVAVAKGPRIVAIISAPLYVQDEVINLTCKAIAFPQPSFH 479  
QY 464 MNVNGTASEDDQDQVRLSTLNVLTPELLETGYECTASNDLGKNSILFLELVNLTTLT 523  
DB 480 WSVNGTTHY-MENQHNASNLTVVSHDLRAGAMCRVSNALGVSE-----KHQLDLDDK 533  
QY 524 PDSMTTGLSTASPTTRANSTSTERKLPPEPSRGVIVAVIYCLIVLAVLGVLEYEL 563  
DB 534 PS-----ESKGIITVAIIYCLIVLAVLGVSTIYELH 563  
QY 584 KKGKLPGRSGKQEI 598  
DB 564 KKGKISGCRSGKQDI 578

RESULT 6  
Q98923 PRELIMINARY; PRT; 504 AA.

AC Q98923;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
DE HEMCAM PRECURSOR.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H.B19OV+; TISSUE=THYMUS;  
RA MEDLINE=97133433; PubMed=8978830;  
RX Valino O., Dunon D., Aissi F., Dangy J.P., McNagny K.M., Imhof B.A.;  
RT "HEMCM, an adhesion molecule expressed by c-kit+ hemopoietic  
RT progenitors.";  
RU J. Cell Biol. 135:1655-1668(1996).  
DR EMBL: Y08854; CAA70079.1; -;  
DR INTERPRO: IPR003006; -;  
DR PFAM: PF00047; 19; 5;  
KW Signal.  
FT SIGNAL 1 35 POTENTIAL.  
FT CHAIN 36 504 POTENTIAL.  
SQ SEQUENCE 504 AA; 55540 MW; ADAE4B94ED4F02E1 CRC64;

Query Match 22.2%; Score 747; DB 13; Length 504;  
Best Local Similarity 35.9%; Pred. No. 3.7e-50;  
Matches 172; Conservative 91; Mismatches 184; Indels 32; Gaps 13;

QY 8 CAFLAACCCCPVAGVGEAPAPPELVEVEVSTALCKGLS--OSGCLSHVDVSVH 66  
DB 22 CPEFLCLLC-----GAAGRLVYMPAVLVELEGSTARLECSFSIPGNSFSISMTFYV 76  
QY 67 KEKRTLIFRRGOG-----QSEPEYEQRLSLQDGAFLALQVTPDE-RIFLCQ 117  
DB 77 RRP-----LRSGEAVRNHNSGVRIDETEYSERLSVGEDKA-LSISKVTRDNARTFIQ 129  
QY 118 --GKRPSOEYRIQLRVYKAPKEPNIOVNPGLIPVNSKEPEVATCVGRNGYP 175  
DB 130 VGADSGVGSERLELYTKIPAPPEITPNSAGIPQSDNMLKIAQCTSENSEFPSPNTTW 189  
QY 176 KNGRPLKEKRVHIOSSQVYESSGLTLOSILKAQLVKEDKQAFCELYNRLPSGNH 235  
DB 190 KNGEPLQEEKTKILTTLVRESNGLYTVSTLFKVTREDRNSLFHCTVHWLOGQRT 249  
QY 236 KESRETVVPEYPTKYLEVER-VGMLKEDGRVEICLADGNPPHFSISKQ--NPSTR 292  
DB 250 KDSRVNVTVYPTHEVELRVATNAGIVKEDDVKLCDDADGNAPAFSEFRELDSMQ 309  
QY 293 EAEETTNDNGVLLEPARKHSGRYCQGLDITMTLSLSEPOELLVNVSDVAV--SP 350  
DB 310 DMTSLADTNDVLMHNVSKSSGLYRCQTLDDMTQHGED-VELVNVYIEGVYKMEP 368  
QY 351 AAPREGSSSLTTCEAESSQDLFEQWLRRETOGVLEGRVQLDHLDRAGGGRVAVS 410  
DB 369 SSP-LHEGDSVRLSCTAHSPVKLDYQW-RDARGKVAEGNQLLTNTLTFETSSNFSCHVK 426  
QY 411 VPSIPCLNTOLVNVAIFGPPMAFKERKYVMKENVNLNLSCEASGHPRTISNVNT 469  
DB 427 ANSVGLEOSKQVAVAKGPRIVAIISAPLYVQDEVINLTCKAIAFPQPSFHWISNGT 485

RESULT 7  
Q9JKB2 PRELIMINARY; PRT; 622 AA.

AC Q9JKB2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE LUTHERAN GLYCOPROTEIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV;  
RA Lee G., Willig T.-N., Parsons S.F., Anstee D.J., Mohandas N.,



RA Chasis J.A.;  
 RT "Mouse Lutheran Glycoprotein Gene";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF246667; AAF61742.1; -  
 SQ SEQUENCE 622 AA; 67543 MW; 3A877738D96f6CD9 CRC64;

Query Match 18.4%; Score 618; DB 11; Length 622;  
 Best Local Similarity 28.5%; Pred. No. 6,1e-40;  
 Matches 191; Conservative 102; Mismatches 275; Indels 102; Gaps 25;

QY 10 FLAACCPCPRVAGVGEAPAEPELVEVEGSTALLKCGLSQSGNLSH--VDMF----- 63  
 DB 15 FLTSG-----YSGAQALHVSVPPEVMEGEOVALDC---TPREHEHVLEMFVLDG 65  
 QY 64 -----SVHKEKRTLIFRVKOGGSEPEYEQRLSLDQRGATLALTOYTPODERITL 115  
 DB 66 TGARHRLASVPEQSEFGLTV-HSLGRVPPEVDSR-----GRLVIKAYOVGDGRDYV 117  
 QY 116 CQGR--RRSQEYRIQLRVYKAPPEPNIOVNPGLIPVNSKEPEFATCGVNGYPIPOVI 173  
 DB 118 CVVAGAAGTSEATSSVAVFATPEDEYSPNKGTLVMDQPAQETATCSSNNGNPVPRT 177  
 QY 174 WYKNGRPLK--EKNRVHIQSQTVESSGLYTLQSLKADLVKEDKDAQFCYCELYRLP 230  
 DB 178 WYRNGQRLLEVMEVNOGXITIRTVREASGLYSLTFLYRLHKDDRDADPHCAHYDLP 237  
 QY 231 SGNHMK-ESREVTVPVPTPKV--WLEVEPV--GMLKEGDRVETRIADGNPPPHSIS 285  
 DB 238 SGQGRDLSTHFRLTLHPTHEVHEFWGSPSTTEGWRGDAVOLLCOGDSSPSPEYSFF 297  
 QY 286 KQNSTREAEETNDNGVLVLEPARKHSGRYEC--OGLDIDMISLSPEOLLVYV 343  
 DB 298 RQO-GTQE-EOLNVNLKGNLTLEGVHRNQSITGCRVEDYADAEVOLYVK-LKLHVAYL 354  
 QY 344 SDVRVSPAPR---QEGSSLTITCEASSQDLEFQWLREBTGOVLERGPVLQHLDKR 399  
 DB 355 DPLELS--APELEFVFLNSSSTVNCASARGLPPTVWTKDSV--TLADGPMLSQSVTF 410  
 QY 400 EAGGTCVAVSPSTPGNLRTQLVNVAIFGPPM-----AKERKRVYKEMVNLSCSEA 454  
 DB 411 DSAGTYCEASTPTVPLLSRTQSFOLIVQGAPELKPNEIMPKSGNSWTEGDEVM-LTCSA 469  
 QY 455 SGHPRPTISMVNGTASBODDQPR--VLSTLVNLVPELLETGVECTASNDLKNITSIL 512  
 DB 470 RGFPEPKLWQSGRDTAEPPFEGRGKSSSLMKVTSALREGVSCASINHGKGVHF 529  
 QY 513 FLELVNLTTLTPDSNTTGLSTASPHTRANSTSTERKLPPEPESRGVIVAVIYCIYL 572  
 DB 530 HFGSV-----APQTAQAGVAVMAVAVSVGLL 555  
 QY 573 AVLGAVALFLYKKKGLKCRSGKQETILPPSRKSELVVEVSKDLPEMGLQ--SSGD 630  
 DB 556 LLVVAALFYCMRRKGRPGCC--RAEKAPAPAREPL--SHSGSERPHTGLMGPSGG 611  
 QY 631 KRAPDQGEK 640  
 DB 612 RGSSEGGFDE 621

RESULT 8  
 Q9R069 PRELIMINARY; PRT; 622 AA.  
 AC Q9R069;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE LUTHERAN GLYCOPROTEIN.  
 GN GPIU OR LU.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Rahuel C., Colin Y., Goossens D., Gane P., El Nemer W., Carron J.P.,  
 RA Le Van Kim C.;  
 RT "Characterization of a mouse laminin receptor gene homologous of the  
 RT human blood group Lutheran gene";  
 RL Immunogenetics 0:0-0(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Parsons S.F., Lee G., Chasis J.A., Tanner M.J.A., Anstee D.J.;  
 RT "Mouse Lutheran glycoprotein";  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF109160; AAF14226.1; -  
 DR EMBL: AF221507; AAF34657.1; -  
 DR INTERPRO: IPR003006; -  
 DR PFM; PF00047; 19; 5.  
 SQ SEQUENCE 622 AA; 67670 MW; 257903F08D47EB4A CRC64;

Query Match 18.3%; Score 614; DB 11; Length 622;  
 Best Local Similarity 28.4%; Pred. No. 1,3e-39;  
 Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FLAACCPCPRVAGVGEAPAEPELVEVEGSTALLKCGLSQSGNLSH--VDMF----- 63  
 DB 15 FLTSG-----YSGAQALHVSVPPEVMEGEOVALDC---TPREHEHVLEMFVLDG 65  
 QY 64 -----SVHKEKRTLIFRVKOGGSEPEYEQRLSLDQRGATLALTOYTPODERITL 115  
 DB 66 TGARHRLASVPEQSEFGLTV-HSLGRVPPEVDSR-----GRLVIKAYOVGDGRDYV 117  
 QY 116 CQGR--RRSQEYRIQLRVYKAPPEPNIOVNPGLIPVNSKEPEFATCGVNGYPIPOVI 173  
 DB 118 CVVAGAAGTSEATSSVAVFATPEDEYSPNKGTLVMDQPAQETATCSSNNGNPVPRT 177  
 QY 174 WYKNGRPLK--EKNRVHIQSQTVESSGLYTLQSLKADLVKEDKDAQFCYCELYRLP 230  
 DB 178 WYRNGQRLLEVMEVNOGXITIRTVREASGLYSLTFLYRLHKDDRDADPHCAHYDLP 237  
 QY 231 SGNHMK-ESREVTVPVPTPKV--WLEVEPV--GMLKEGDRVETRIADGNPPPHSIS 285  
 DB 238 SGQGRDLSTHFRLTLHPTHEVHEFWGSPSTTEGWRGDAVOLLCOGDSSPSPEYSFF 297  
 QY 286 KQNSTREAEETNDNGVLVLEPARKHSGRYEC--OGLDIDMISLSPEOLLVYV 343  
 DB 298 RQO-GTQE-EOLNVNLKGNLTLEGVHRNQSITGCRVEDYADAEVOLYVK-LKLHVAYL 354  
 QY 344 SDVRVSPAPR---QEGSSLTITCEASSQDLEFQWLREBTGOVLERGPVLQHLDKR 399  
 DB 355 DPLELS--VPELEFVFLNSSSTVNCASARGLPPTVWTKDSV--TLADGPMLSQSVTF 410  
 QY 400 EAGGTCVAVSPSTPGNLRTQLVNVAIFGPPM-----AKERKRVYKEMVNLSCSEA 454  
 DB 411 DSAGTYCEASTPTVPLLSRTQSFOLIVQGAPELKPNEIMPKSGNSWTEGDEVM-LTCSA 469  
 QY 455 SGHPRPTISMVNGTASBODDQPR--VLSTLVNLVPELLETGVECTASNDLKNITSIL 512  
 DB 470 RGFPEPKLWQSGRDTAEPPFEGRGKSSSLMKVTSALREGVSCASINHGKGVHF 529  
 QY 513 FLELVNLTTLTPDSNTTGLSTASPHTRANSTSTERKLPPEPESRGVIVAVIYCIYL 572  
 DB 530 HFGSV-----APQTAQAGVAVMAVAVSVGLL 555  
 QY 573 AVLGAVALFLYKKKGLKCRSGKQETILPPSRKSELVVEVSKDLPEMGLQ--SSGD 630  
 DB 556 LLVVAALFYCMRRKGRPGCC--RAEKAPAPAREPL--SHSGSERPHTGLMGPSGG 611  
 QY 631 KRAPDQGEK 640  
 DB 612 RGSSEGGFDE 621

RESULT 9

035112 PRELIMINARY; PRT; 583 AA.  
 ID 035112: 055172;  
 AC 035112: 055172;  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE CD166 ANTIGEN PRECURSOR (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE)  
 DE (ALCAM) (MEMB PROTEIN) (HB2) (KG-CAM).  
 GN ALCAM.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 -OX NCBI\_TaxID=10116;  
 RN RN  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=97347476; PubMed=9201982;  
 RA Matsumoto A., Mitchell A., Kurata H., Pyle L., Kondo K., Itakura H.,  
 RA Flidge N.;  
 RT "Cloning and characterization of HB2, a candidate high density  
 RT lipoprotein receptor. Sequence homology with members of the  
 RT immunoglobulin superfamily of membrane proteins.";  
 RL J. Biol. Chem. 272:16778-16782(1997).  
 RN [2]  
 RP SEQUENCE OF 270-361 FROM N.A.  
 RC TISSUE=MESENCHYME;  
 RX MEDLINE=98215351; PubMed=9556065;  
 RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,  
 RA Zala J., Barry F.P.;  
 RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated  
 RT leukocyte cell adhesion molecule and is involved in osteogenic  
 RT differentiation.";  
 RL J. Bone Miner. Res. 13:655-663(1998).  
 RN [3]  
 RP SEQUENCE OF 29-38.  
 RX MEDLINE=94272944; PubMed=8004458;  
 RA Peduzzi J.D., Irwin M.H., Geisert E.E. Jr.;  
 RT "Distribution and characteristics of a 90 kDa protein, KG-CAM, in the  
 RT rat CNS.";  
 RL Brain Res. 640:296-307(1994).  
 CC CC  
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN  
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC  
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO  
 CC ACTIVATED LEUKOCYTES, AS WELL AS IN INTERACTIONS BETWEEN CELLS OF  
 CC THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN  
 CC IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL  
 CC TISSUES (BY SIMILARITY).  
 CC CC  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC CC  
 CC -1- TISSUE SPECIFICITY: STRONGEST EXPRESSION IN THE LUNG, THEN BRAIN,  
 CC LIVER, AND KIDNEY. PRESENT IN THE SOMATOSENSORY SYSTEM, BASAL  
 CC GANGLIA, CORTEX, OLFACTORY SYSTEM, AND CIRCUMVENTRICULAR ORGANS.  
 CC CC  
 CC -1- PTM: THE N-TERMINAL IS BLOCKED.  
 CC CC  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3  
 CC C2-LIKE AND 2 V-LIKE DOMAINS.  
 CC EMBL; AB008538; BAA23279.1; -;  
 DR EMBL; Y13241; CA73693.1; -;  
 DR EMBL; Y13240; CA73692.1; -;  
 DR HSSP; Q13740; IKJC.  
 DR INTERPRO: IPR003006; -;  
 DR PFAM; PF00047; 19; 5.  
 KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;  
 KW Signal; Antigen.  
 FT SIGNAL 1 27  
 FT CHAIN 28 583  
 FT DOMAIN 28 527  
 FT TRANSMEM 528 548  
 FT DOMAIN 549 583  
 FT DOMAIN 36 120  
 FT DOMAIN 150 227  
 FT DOMAIN 263 320  
 FT DOMAIN 347 399  
 FT DOMAIN 428 492  
 FT DISULFID 43 113  
 FT DISULFID 157 220  
 FT POTENTIAL.  
 FT CD166 ANTIGEN.  
 FT EXTRACELLULAR (POTENTIAL).  
 FT POTENTIAL.  
 FT CYTOPLASMIC (POTENTIAL).  
 FT IG-LIKE V-TYPE DOMAIN.  
 FT IG-LIKE V-TYPE DOMAIN.  
 FT IG-LIKE C2-TYPE DOMAIN.  
 FT IG-LIKE C2-TYPE DOMAIN.  
 FT IG-LIKE C2-TYPE DOMAIN.  
 FT POTENTIAL.  
 FT POTENTIAL.

FT DISULFID 270 313 POTENTIAL.  
 FT DISULFID 354 392 POTENTIAL.  
 FT DISULFID 435 485 POTENTIAL.  
 FT CONFLICT 339 339 S -> G (IN REF. 2).  
 FT CARBOHYD 95 95 POTENTIAL.  
 FT CARBOHYD 167 167 POTENTIAL.  
 FT CARBOHYD 265 265 POTENTIAL.  
 FT CARBOHYD 306 306 POTENTIAL.  
 FT CARBOHYD 361 361 POTENTIAL.  
 FT CARBOHYD 457 457 POTENTIAL.  
 FT CARBOHYD 480 480 POTENTIAL.  
 FT CARBOHYD 499 499 POTENTIAL.  
 SQ SEQUENCE 583 AA; 65021 MW; D36B73854F5D61E CRC64;  
 Query Match 14.9%; Score 499.5; DB 11; Length 583;  
 Best Local Similarity 24.6%; Pred. No. 1e-30;  
 Matches 153; Conservative 122; Mismatches 273; Indels 73; Gaps 25;  
 QY 5 RLVCAPFLAACCCCPRVAGVGEAQPAPPELVVEVSTALTKGLSGSGNLSHVDFPS 64  
 DB 10 RLVCCLLSAVALRPGLGWY-----TVNSAVGDTLVMPCRIDVQ-NLMFGKMY 58  
 QY 65 VHKERTLIRVRGQGS-----EPGEYDRLSLDRGATLATQYTPDDEIFLCQ-GK 119  
 DB 59 EKPDSVVFIAFRSTKKSQYDVPDEYKDRLSLSE-NYTLSSINNAKISDEKRFVCLVT 117  
 QY 120 RPRSGEIRIDLVYKAPPEINIOVNPILGIPNKEPEVATCYGRGYPPIQVYWKGR 179  
 DB 118 EDNVFEAPLVKVKFKQSKPEI-VNRAAF-LETEQLKGLDCCISRSYDPGNITWYRNRK 175  
 QY 180 PLKEKNRVHIQSSQYVE-SSGLYTLQSLKAQVLKEDKQAPCYELNRLPSGNHKS 238  
 DB 176 VLQPVDESEVILFKKEIDPQTLTYTSSLEYTTSDIMPTGCVTYGSGQKTIIS 235  
 QY 239 REYTVPVFYETEKVWLEV-EPVGLKEGDRVELRCLADGNPPH---FSISKONPSTREA 294  
 DB 236 EQALFEDIYPTQEVTLQVLPKKNAIKEGDMITLQCLNGNPPPEEFMYLPQGAEGIRSS 295  
 QY 295 EEEFTNNGVLVEPARKHSRGYECQGLDLDIMISLSPQGLVNVYSQVAVPSAAV- 353  
 DB 296 NYTLVD-----VRRATGDKCSLDORNMMASTT---ITVHYL-DLSLNSGEV 342  
 QY 354 EROGSSLTITCEAESSQDLDFQMLRETFQVLERGVLDLHDKREAGGYRCVAVSPS 413  
 DB 343 TKQIGDPLPVSCITISARNATVWMKNIR--LRSP--SFSSLHQDAGNYCETALAE 398  
 QY 414 IPGNRTQLVNVAIFGPPMAFKERKVKYKENVNLNLSCEASGHPPTISMNVNGTAS-- 471  
 DB 399 VEGIKKRESLTLVEGRK--QIMTKTKTPSGSLKTLICVEGFPPALQWITVGSQSVI 456  
 QY 472 -EDQDDQ--RYLSTNLVTPPELLETGEVCTASNDLKNSTILFLELVNLTTLTPDSN 527  
 DB 457 NQTEESPINGRYIS--KIISPEENV--LTCTAENQLEFTVSLVNSAIST---PE-- 507  
 QY 528 TTGLSTSTASPHTRANSTSTERKLEPEPSRGVIAVAVICILAVLAVGLVFLY-KRG 586  
 DB 508 -----HDEADDIDENR-EKVNDQAKLIIYIVGILLALVAGVYVLMKKMS 554  
 QY 587 KLPGRSGKQETILPSRKSE 607  
 DB 555 KTSKSHVKNKDLGNMEENKLE 575  
 RESULT 10  
 ID 046634 PRELIMINARY; PRT; 521 AA.  
 AC 046634;  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)  
 DE CD166 ANTIGEN (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE) (ALCAM)  
 DE (FRAGMENT).

GN ALCAM.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
ON NCBI\_Taxid=9615;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MESENCHYME;  
RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,  
RT Zaia J., Barry F.P.;  
RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated  
RT leukocyte cell adhesion molecule and is involved in osteogenic  
RT differentiation."  
RL J. Bone Miner. Res. 13:655-663(1998).  
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN  
CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC  
CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO  
CC ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF  
CC THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN  
CC IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL  
CC TISSUES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3  
CC C2-LIKE AND 2 V-LIKE DOMAINS.  
DR EMBL: Y13242; CAI73694.1; -.  
DR HSSP: Q13740; IKJC.  
DR INTERPRO: IPR003006; -.  
DR PFAM: PF00047; 19; 4.  
KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN 1 466 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 467 487 POTENTIAL.  
FT DOMAIN 488 521 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN <1 58 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 88 165 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 201 259 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 285 337 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 366 430 IG-LIKE C2-TYPE DOMAIN.  
FT DISULFID <1 51 POTENTIAL.  
FT DISULFID 95 158 POTENTIAL.  
FT DISULFID 208 251 POTENTIAL.  
FT DISULFID 292 330 POTENTIAL.  
FT DISULFID 373 423 POTENTIAL.  
FT CARBOHYD 33 33 POTENTIAL.  
FT CARBOHYD 105 105 POTENTIAL.  
FT CARBOHYD 244 244 POTENTIAL.  
FT CARBOHYD 299 299 POTENTIAL.  
FT CARBOHYD 395 395 POTENTIAL.  
FT CARBOHYD 418 418 POTENTIAL.  
FT CARBOHYD 437 437 POTENTIAL.  
SQ SEQUENCE 521 AA; 38106 MW; A3616E9A2429E7B6 CRC64;

Query Match 14.4%; Score 485; DB 6; Length 521;  
Best Local Similarity 26.4%; Pred. No. 1.2e-29;  
Matches 143; Conservative 107; Mismatches 220; Indels 72; Gaps 22;

QY 314 HSGRYECGLDLDIMISLSEPELVNYSVAVSPAP--EROEGSLLTLCGEASSOD 372  
DB 245 ANGDYCKSLIDKSKMIASTA---ITVHYL-DLSINBGEVTKOIGALPVSCTISASRN 299  
QY 373 LEFOWLREFTGVLERGVVLQHLDKREAGGGYCVASVPSPIDNLTOLVNAIFGPPW 432  
DB 300 ATVVMKDNIR--LRSSP--SFSSIQYODAGNYCETALQVEBKKRESLTLLVEKP- 354  
QY 433 NAFKRRKVVVENWVLNLSCEASGHPPTISWNVNGTAS---EDQDPQ---RYLSLNV 486  
DB 355 -QIKMTKTKDPSGLSKTITICHEGFPKPAIOWTITGSGSVYNQTEESPYINGRYST--I 411  
QY 487 LVTPPELEGVGCTASNDLKNSTLTFLELVNLTLPDPSNTTGLSTASPTPTANST 546  
DB 412 INSPEENVT-LTCTIENDLERNTLVNSASTI-----PE-----HDEDEI 452  
QY 547 STERKLPEPESRGVIVAVICILVAVGAVLYELY-KKGLPCRRSGKQELTLPSPRK 605  
DB 453 SPENR-EQVNHKATLIVGLVRLHGLVAGVYVLYKSKTASKHYNKDGLNLEKMK 511  
QY 606 SE 607  
DB 512 LE 513

RESULT 11  
ID 046651 PRELIMINARY; PRT; 521 AA.  
AC 046651;  
DT 01-JUN-1998 (TREMBLrel. 06, created)  
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)  
DE CD166 ANTIGEN (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE) (ALCAM)  
DE (FRAGMENT).  
DE GN  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
ON NCBI\_Taxid=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=MESENCHYME;  
RX MEDLINE=98215351; PubMed=9556065;  
RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,  
RT Zaia J., Barry F.P.;  
RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated  
RT leukocyte cell adhesion molecule and is involved in osteogenic  
RT differentiation."  
RL J. Bone Miner. Res. 13:655-663(1998).  
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN  
CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC  
CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO  
CC ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF  
CC THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN  
CC IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL  
CC TISSUES (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3  
CC C2-LIKE AND 2 V-LIKE DOMAINS.  
DR EMBL: Y13243; CAI73695.1; -.  
DR HSSP: Q13740; IKJC.  
DR INTERPRO: IPR003006; -.  
DR PFAM: PF00047; 19; 4.  
KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN 1 466 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 467 487 POTENTIAL.  
FT DOMAIN 488 521 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN <1 58 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 88 165 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 201 259 IG-LIKE C2-TYPE DOMAIN.  
FT DOMAIN 285 337 IG-LIKE C2-TYPE DOMAIN.

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Alais S., Allioti N., Vigneron P., Dunon D.;
RT "HEMCAM is the avian homologue of MCAM/CD146.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ297451; CAB97363.1; -.
FT NON_TER
SQ SEQUENCE 88 AA; 9588 MW; AAF4AD86DE0D3B CRC64;

Query Match 12.1%; Score 407; DB 11; Length 88;
Best Local Similarity 90.9%; Pred. No. 1, 1e-24;
Matches 80; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 559 GYIVAVIYCIILVALVGLAVLYFLYKKGKILPCRSCKOETILPPSKSELYEVKSDKLP 618
DB 1 GYIVAVIYCIILVALVGLAVLYFLYKKGKILPCRSCKOETILPPSKSELYEVKSDKLP 60
QY 619 EEMGLQGSSGDKRAPGDGGEKYIDLRH 646
DB 61 EEMALLQGSNGDKRAPGDGGEKYIDLRH 88

RESULT 13
ID 019129 PRELIMINARY; PRT; 86 AA.
AC 019129;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE MELANOMA CELL ADHESION MOLECULE (FRAGMENT).
GN MCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RA Moore S.S., Byrne K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89328; AAB6511.1; -.
FT NON_TER
SQ SEQUENCE 86 AA; 9486 MW; F109345FDB2B3555 CRC64;

Query Match 10.8%; Score 362; DB 6; Length 86;
Best Local Similarity 93.3%; Pred. No. 3, 5e-21;
Matches 70; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 572 LAVIGAVILFYLKRRKILPCRSRSGKQETILPPSRKSELYEVKSDKLPPEMGLQGSSGDK 631
DB 12 LGLIGAVILFYLFYKRRKILPCGRSGKQETILPPSRKSELYEVKSDKLPPEMGLQGSSGDK 71
QY 632 RAPGDGGEKYIDLRH 646
DB 72 RAPGDGGEKYIDLRH 86

RESULT 14
ID 09M208 PRELIMINARY; PRT; 355 AA.
AC 09M208;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE LUTHERAN GLYCOPROTEIN (FRAGMENT).
GN LU.
OS Bos taurus (Bovine).

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 CC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Boado R.U., Li J.Y., Pardridge W.M.;  
 RT "Selective Lutheran glycoprotein gene expression at the blood-brain  
 barrier in normal brain and in human brain tumors."  
 RL J. Cereb. Blood Flow Metab. 0:0-0(2000).  
 FT EMBL: AF270512; AAF81749.1; -  
 SO NON\_TER

Query Match 10.7%; Score 361; DB 6; Length 355;  
 Best Local Similarity 28.6%; Pred. No. 3.3e-20;  
 Matches 114; Conservative 63; Mismatches 155; Indels 66; Gaps 15;

QY 260 GMLKGGVETIRCLADGNPPPHFSI---SKONPSTREAEETINDGVLEPARKHS 315  
 D 6 GMLKGGVETIRCLADGNPPPHFSI---SKONPSTREAEETINDGVLEPARKHS 315  
 QY 316 GRYCQGLDPTMISTLSEPOELVYVSDVAVSPAPRQE---GSSLTTCCEASSQ 371  
 D 60 GRYCQGLDPTMISTLSEPOELVYVSDVAVSPAPRQE---GSSLTTCCEASSQ 371  
 QY 372 DLEFQWIREETGVLENGVPILOHDLKREAGGRCVAVSPISPTGLNRTQVNAVATGCP 431  
 D 118 TPLTYWTKDAS--PMGEDPTLSHVSFDSAGYTCAYMPRIPLLSRTSRFLVVGTP 431  
 QY 432 WMAEKREK---VWVENVNLNLSCEASGPHPTISMN-VNCTASPODDPOR-VLSTL 484  
 D 176 ELKAKETOPRAEGSWTEGDEV-FLICVARGYPRPKLWLSGSPTEPRAGCGGWSSSL 234  
 QY 485 NVLTPBELLETGVECTASNDLGKNTSLFLELVNLTLPDSNTTGLSTASPHTRAN 544  
 D 235 TLKTSALSDQGVSCASNPLG-----VTHHVFHFQVAPQT--- 271  
 QY 545 STERKLPPEESGVYVAVICILVAVLAVLFLYKKGKLPCKRSKQKQITLPPSR 604  
 D 272 -----SOAGVAVMVAISVALLLVAVVFCMRKGRPGCCQWG--EKGSPPG 318  
 QY 605 KSELVEVKSCLKPEENGILQSS--GDKRAPDQGEK 640  
 D 319 EPKL--SHSGSQREQTGILMGASAGAKHSGGFQDE 354

RESULT 15  
 076518 PRELIMINARY; PRT; 5198 AA.  
 AC 076518; Q10036;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE HEMICENTIN PRECURSOR.  
 GN Hm-4 OR F15G9.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OC NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA Vogel B.E., Hedgecock E.M.;  
 RT "Hemicentin is Required for Hemidesmosome Mediated Cell Adhesion and  
 RT Germ-Line Chromosome Segregation in C. elegans."  
 RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA Sulston J., Kershaw J.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.

CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, ISOFORM F15G9.4A AND ISOFORM  
 CC F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 DR EMBL: AF074901; AAC26792.1; -  
 DR EMBL: 247068; CAAB7336.1; -  
 DR EMBL: 247070; CAAB7336.1; JOINED.  
 DR EMBL: 247070; CAAB7345.1; -  
 DR EMBL: 247068; CAAB7345.1; JOINED.  
 DR EMBL: 247068; CAAB7335.1; -  
 DR EMBL: 247070; CAAB7335.1; JOINED.  
 DR EMBL: 247070; CAAB7344.1; -  
 DR EMBL: 247068; CAAB7344.1; JOINED.  
 DR HSSP: P35555; IEMN.  
 DR WORMPEP: F15G9.4A; CE18595.  
 DR WORMPEP: F15G9.4B; CE18596.  
 DR INTERPRO: IPR000152; -  
 DR INTERPRO: IPR000515; -  
 DR INTERPRO: IPR000561; -  
 DR INTERPRO: IPR001881; -  
 DR INTERPRO: IPR003006; -  
 DR PFAM: PF00047; 19; 47.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS00402; BPD\_TRANSF\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN\_1.  
 DR PROSITE: PS01187; EGF\_CA\_2.  
 DR Signal: Glycoprotein; EGF-like domain; Alternative splicing.  
 FT CHAIN 1 24 POTENTIAL.  
 FT VARSPIC 25 5198 HEMICENTIN.  
 FT MISSING (IN ISOFORM F15G9.4A).  
 SO SEQUENCE 5198 AA; 570816 MW; DA8511FE258D37B CRC64;

Query Match 8.4%; Score 281; DB 5; Length 5198;  
 Best Local Similarity 22.1%; Pred. No. 3e-12;  
 Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAEOPAPLVEVEVSTAL-----LKGLSQSQGNLSHVMEFVHKKRLIRVRQG- 79  
 D 1357 DVGEPIILPSTQTNNTAVAGRVELCKYEVNSPP--ASVTF-----RGT 1401  
 QY 80 -GGSEPEGEORLSLDGATLALQVTPDERLFLCGKRRPSQ--EYRIQLRYAPE 137  
 D 1402 AIGDTTKG-----YVESDGTFLVQSASVEDATITCKASNPAGKAAQVYVNASPD 1455  
 QY 138 --EPMIYVPLGIPVNSKEPEVATCVGRNGYPIPOVWYKNGRL-----KEEK 186  
 D 1456 IDPPVYTOE---STKESHPESLYCVPFSN--PLQIISWYLNDRPLIDDKTSWKTSDKR 1510  
 QY 187 RVHIQSOTVESSGLYTLQSLIKAQLVKEDKDAQCYCELNRLPSGNMKESR-EVTVPV 245  
 D 1511 KIHVKRAK-IIDSGYKKVA-----RNAAGBSKSFQVEVIVPL 1548  
 QY 246 EYPTPKVWLEVPVGMKEDGVEIRCLADGNPPPHFSI---SKONPSTREAEETIND 301  
 D 1549 NLDESKYKKV---FAKEGEVTLGCPVSGFPVQIMWVDGIVYERGGKRYKATLSND 1604  
 QY 302 NCVLLEPARKHSGRYECOG-----LDPTMISTLSEPOELVYVSDVAVSPAPR 355  
 D 1605 GLTLFDFSVAKQEBNHYCAVQSKNIIIDVLSVLAVP---IVGEDDNLEVF----- 1655  
 QY 356 QEGSSLTLTC--EAESSDLEFQWL-----REETGVLENGVPILOHDLKREAGGGR 406  
 D 1656 -LGKDLSLSCDLQTFESDQKTFVWSINGSESDRDQVQIPDGDHRLVITPAKPNNNKYM 1714  
 QY 407 CVASVPSIPGLNRQVLVAV---AIFGPPWMAFKERKVVKNVNLNLSCEASGPHPTIS 463  
 D 1715 CRVT-NSAGKAERLTLLDLEPPVFEV--VEFANOKLIGNNPLT-LQCVGTGNPKPTVI 1770  
 QY 464 WNVNSTAGEODDQPVRLSTNLVLTPELTG---VECTASNDLGKNTSLFLELVNLT 519  
 D 1771 WKIDGNDVCKSWLDEDSILRTI-----EKL-TGSAQISCTAENKAGTASRDFEIONIAA 1825  
 QY 520 TYLPPDSNTTT 530

Mon Mar 26 09:22:40 2001

us-09-653-961-2.rspt

Page 10

Db 1826 PTFKNEGDET 1836

Search completed: March 23, 2001, 12:32:35  
Job time: 113 sec

---